

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 20:53:41 ; Search time 53 Seconds
(without alignments)
8275.405 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattgataaca.....tttagacttacgaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PT05_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	103.2	6.1	7218	1	US-08-232-463-14
C 2	76	4.5	19124	2	US-08-487-826B-13
C 3	71.2	4.2	152331	3	US-09-128-155-16
C 4	71.2	4.2	176373	3	US-09-128-155-17
C 5	62.6	3.7	152331	3	US-09-128-155-16
C 6	58	3.4	6124	4	US-08-213-419B-3
C 7	56.8	3.3	289	4	US-09-007-005-17
C 8	56.8	3.3	289	4	US-09-244-796-17
C 9	55.6	3.3	291	1	US-07-922-723A-7
C 10	55.6	3.3	291	1	US-07-799-828C-7
C 11	55.6	3.3	291	1	US-08-074-275-7
C 12	55.6	3.3	291	1	US-08-480-366-7
C 13	55.6	3.3	291	2	US-07-952-277A-7
C 14	55.6	3.3	19124	2	US-08-487-826B-13
C 15	55.6	3.3	44453	3	US-09-146-053-5
C 16	54.2	3.2	5852	1	US-07-867-106-2
C 17	53.4	3.1	1956	4	US-08-559-896B-1
C 18	52.8	3.1	7379	4	US-09-341-587-5
C 19	52.4	3.1	1298	4	US-08-971-089-1
C 20	51.8	3.0	454	2	US-08-623-906A-6
C 21	51.6	3.0	72604	4	US-09-268-992-7
C 22	51.6	3.0	72604	4	US-09-657-474-7
C 23	50.4	3.0	1559	4	US-09-019-095A-7
C 24	50.2	3.0	2791	4	US-09-570-367C-1
C 25	50	2.9	1298	3	US-08-948-705-3
C 26	49.6	2.9	417	4	US-08-559-896B-3
C 27	49.6	2.9	1859	3	US-08-691-563C-46

28	49.6	2.9	45546	4	US-09-146-053-6	Sequence 6, Appli
29	49.2	2.9	2223	1	US-08-257-073-4	Sequence 4, Appli
30	49	2.9	2040	1	US-08-599-252-103	Sequence 103, App
31	49	2.9	2040	5	PCT-US96-06352-103	Sequence 103, App
32	49	2.9	2040	5	PCT-US96-06583-103	Sequence 103, App
33	48.6	2.9	12141	4	US-09-488-671-10	Sequence 10, Appl
C 34	48.2	2.8	372	4	US-09-018-584A-13	Sequence 13, Appl
C 35	47.8	2.8	658	4	US-08-998-416-595	Sequence 595, App
C 36	47.8	2.8	72604	4	US-09-268-992-7	Sequence 7, Appli
C 37	47.8	2.8	72604	4	US-09-657-474-7	Sequence 7, Appli
C 38	47.2	2.8	2447	2	US-09-014-969-14	Sequence 14, Appli
C 39	46.8	2.8	2030	4	US-09-512-342-1	Sequence 1, Appli
C 40	46.6	2.7	240	1	US-08-628-417-6	Sequence 6, Appli
C 41	46.6	2.7	1214	4	US-08-817-913-13	Sequence 13, Appli
C 42	46.6	2.7	1232	4	US-08-817-913-14	Sequence 14, Appli
C 43	46.6	2.7	1352	4	US-08-817-913-15	Sequence 15, Appli
C 44	46.6	2.7	1734	4	US-08-817-913-16	Sequence 16, Appli
C 45	46.6	2.7	1920	4	US-08-817-913-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzpgt-fls
; US-08-232-463-14

Query Match 6.1%; Score 103.2; DB 1; Length 7218;

Db 33 RARCRARURURARCRARURGRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 92
QY 255 TGAGGTGAACAAAATAAGAGATATGATCATATATTTATGAGAGGTGGTGAAGATTAT 314
Db 93 RNR 152
QY 315 TTTAGGAGGGAGAGAGAAATAGAAAAAGAAATGACATGTTGAATCTGAAGAAGATGA 374
Db 153 RNR 212
QY 375 ATTGTTGTTAAAGATGAAGAGAGAAAGAACTCCATGGCTTAAAGTCTCGTAAAGAGATG 434
Db 213 RNRNRRCRARGRURGRURGRURARARCRURURGRURGRURGRURGRURGRUR 272
QY 435 AAAAAAGAAACAAA 448
Db 273 AAAAAAGAAACAAA 286

RESULT 9

US-07-922-723A-7

; Sequence 7, Application US/07922723A

; Patent No. 5369004

; GENERAL INFORMATION:

; APPLICANT: Drs. Mihael H. Polymeropoulos

; APPLICANT: and Carl R. Merrill

; TITLE OF INVENTION: FIVE HIGHLY INFORMATIVE

; TITLE OF INVENTION: REPEAT POLYMORPHIC DNA MARKERS

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lowe, Price, LeBlanc & Becker

; STREET: Suite 300, 99 Canal Center Plaza

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: DOS Text File

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/922,723A

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: D.J. Mills

; REGISTRATION NUMBER: 34506

; REFERENCE/DOCKET NUMBER: 717081B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703 684 1111

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 291

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-07-922-723A-7

Query Match

Best Local Similarity 3.3%; Score 55.6; DB 1; Length 291;

Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 237 AAAAAAGAGAGAACATGTAGGTGAACAAAATAAGAGATATGATGATATATTTATG 296
Db 33 AGAAAAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAGAA 92
QY 297 AGAGGTGTTGAAGATTATTTTAGGAGGGGAGAGAGAAATAGAAAAAGAAATGACATGG 356
Db 93 AGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152

QY 357 TGAATCTGAAGAAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAA 416
Db 153 AGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 212
QY 417 AGTCTCTGTAAGACAGATGAAAAAGAAACAAAGAAAGAGGAGAAAGAAAGAGA 466
Db 213 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 262

RESULT 10

US-07-799-828C-7

; Sequence 7, Application US/07799828C

; Patent No. 5378602

; GENERAL INFORMATION:

; APPLICANT: Drs. Carl R. Merrill and

; APPLICANT: Mihael H. Polymeropoulos

; TITLE OF INVENTION: TWENTY SEVEN HIGHLY INFORMATIVE

; TITLE OF INVENTION: MICROSAATELLITE REPEAT

; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS

; NUMBER OF SEQUENCES: 63

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lowe, Price, LeBlanc & Becker

; STREET: Suite 300, 99 Canal Center Plaza

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: DOS Text File

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/799,828C

; FILING DATE: 19911127

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: D.J. Mills

; REGISTRATION NUMBER: 34,506

; REFERENCE/DOCKET NUMBER: 717081A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703 684 1111

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 291

; TYPE: NUCLEIC ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-07-799-828C-7

Query Match 3.3%; Score 55.6; DB 1; Length 291;

Best Local Similarity 52.6%; Pred. No. 0.00065;

Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 237 AAAAAAGAGAGAACATGTAGGTGAACAAAATAAGAGATATGATGATATATTTATG 296
Db 33 AGAAAAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAGAA 92
QY 297 AGAGGTGTTGAAGATTATTTTAGGAGGGGAGAGAGAAATAGAAAAAGAAATGACATGG 356
Db 93 AGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152
QY 357 TGAATCTGAAGAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTAA 416
Db 153 AGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 212
QY 417 AGTCTCTGTAAGACAGATGAAAAAGAAACAAAGAAAGAGGAGAAAGAAAGAGA 466
Db 213 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 262

RESULT 11

US-08-074-275-7
; Sequence 7, Application US/08074275
; Patent No. 5468610
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,275
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/707,501
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 717081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-074-275-7
Query Match 3.3%; Score 55.6; DB 1: Length 291;
Best Local Similarity 52.6%; Pred. No. 0.00065;
Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 237 AAAAAGAAAGAGAGAACATGTTAGGTGAACAAAATAAAGAGATATGATATATTTTATG 296
Db 33 AGAAGAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAGAAAGGA 92
QY 297 AGAGGTGGTGAAGATTATTTTAAAGAGGGGAGAGAGAAATAGAAAAGAAAATGACATGG 356
Db 93 AGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152
QY 357 TGAATCTGAAGAGATGATTTGTTAAAGATGAAGAGAGAGAGAGAACTCCATGGCTAA 416
Db 153 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 212
QY 417 AGTCTCGTAAAGAGATGAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 466
Db 213 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 262
RESULT 12
US-08-480-366-7
; Sequence 7, Application US/08480366
; Patent No. 5721100
; GENERAL INFORMATION:
; APPLICANT: Drs. Carl R. Merrill and
; APPLICANT: Michael H. Polymeropoulos
; TITLE OF INVENTION: THREE HIGHLY INFORMATIVE REPEAT
; TITLE OF INVENTION: POLYMORPHIC DNA MARKERS

; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: DOS Text File
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,366
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: J.G. Mullins
; REGISTRATION NUMBER: 33073
; REFERENCE/DOCKET NUMBER: 717081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703 684 1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-480-366-7
Query Match 3.3%; Score 55.6; DB 1: Length 291;
Best Local Similarity 52.6%; Pred. No. 0.00065;
Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 237 AAAAAGAAAGAGAGAACATGTTAGGTGAACAAAATAAAGAGATATGATATATTTTATG 296
Db 33 AGAAGAAGAAAGAGACAAAGAGAGTTAGAAAGAAAGAAAGAGAGAGAGAGAGAAAGGA 92
QY 297 AGAGGTGGTGAAGATTATTTTAAAGAGGGGAGAGAGAAATAGAAAAGAAAATGACATGG 356
Db 93 AGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 152
QY 357 TGAATCTGAAGAGATGATTTGTTAAAGATGAAGAGAGAGAGAGAACTCCATGGCTAA 416
Db 153 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 212
QY 417 AGTCTCGTAAAGAGATGAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 466
Db 213 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 262
RESULT 13
US-07-952-277A-7
; Sequence 7, Application US/07952277A
; Patent No. 5861504
; GENERAL INFORMATION:
; APPLICANT: Drs. Michael H. Polymeropoulos
; APPLICANT: and Carl R. Merrill
; TITLE OF INVENTION: ELEVEN HIGHLY INFORMATIVE
; TITLE OF INVENTION: REPEAT POLYMORPHIC DNA MARKERS
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, LeBlanc & Becker
; STREET: Suite 300, 99 Canal Center Plaza
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

Query Match 3.3%; Score 55.6; DB 2; Length 19124;
Best Local Similarity 48.4%; Pred. No. 0.0021;
Matches 154; Conservative 0; Mismatches 164; Indels 0;

Query Match 3.3%; Score 55.6; DB 2; Length 291;
Best Local Similarity 52.6%; Pred. No. 0.00065;
Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY	91	AACATGTTTGGTCTTAAAAATACATGCATTTTAAAAATCAGACATTTTCTTTAAAAATCAAA	150
Db	18071	CACATATATAAATACACATTTAAATAATATATTATAAATATCACAATAAATAATAA	18012
QY	151	TCTAATCTTTATATCAACAGCAGCATTTGACGGAAAATTCAGGTAAA-AGAGAAAATAAAG	210
Db	18011	TTAATACTGGAAAAAATGAAAAATCGAAANTAGAGGATATATATATAGAAAAAAGAG	17952
QY	211	AATGAGAGATAGAGAGATTTCTTATGGAAAAAGAAAGAGAGAACATG-ASGTCACACAAAAT	270
Db	17951	AATTATATATATATATATATATATATATATACGAAAAAATGTACAGA-AAGAAAACAAAT	17992
QY	271	AAGAGATATGATGATATATTTTTATGAGAGGTGGTGAAGATTATTTAGGAGAGGGAGAG	330
Db	17891	AACCAATATATACATTAATTTAATTAATCAATATAGATACAAATATTATCTAAAAAATAAA	17832

```

QY      331 AGAAATAGAAAAAGAAA 348
       I   I   I   I   I   I
Db 17831 ATACATATATATATAA 17814

RESULT 15
US-09-146-053-5
; Sequence 5, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146.053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 44453
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-5

```

Query Match 3.3%; Score 55.6; DB 4; Length 44453;
Best Local Similarity 50.0%; Pred. No. 0.0026;
Matches 139; Conservative 0; Mismatches 139; Idels 0;

```

| ||| ||| | | ||| || ||| ||| ||| | | ||| ||| ||| |||
Db 20872 TGAAGGAAAGGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 20931
QY 253 CATCTAGGTGAACAAATAAAGAGATATGATATATATTTTATGAGAGGTGGTGAAGATT 312
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20932 AGAAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 20991
QY 313 ATTTTAGGAGAGGAGAGAGAGAAATAGAAAAAGAAAAATGACATGGTGAATCTGAAAGAAAGAT 372
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20992 AAGAAAAGAGAGAAAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAAACA 21051
QY 373 GAATTGTGTTAAGATGAAGAGAGAAAGAACTCCATGGCTAAAGTCTCGTAAAGAAAGA 432
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21052 GAGAAAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 21111
QY 433 TGAAGAAAGAAACAAAGAGAGAAAGAAAGAGAGAAAGG 470
| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21112 AAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21149
```

Search completed: November 28, 2002, 23:17:15
Job time : 816 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 20:43:06 ; Search time 268 Seconds

(without alignments)
14285.064 Million cell updates/sec

Title: US-09-733-685-3

Perfect score: 1700

Sequence: 1 tgggtttttattgataa.....tttaagactgaacatat 1700

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.8	4.5	7657	22 AAS45477	Chemically pretrea
2	76.8	4.5	7657	24 ABL34022	Human immune syste
3	76.8	4.5	14006	24 ABL33958	Human immune syste
4	76	4.5	19124	18 AAT72882	Plasmodium var-7 g
5	76	4.5	19124	21 AAZ98287	Plasmodium var-7 p
6	74.8	4.4	9539	22 AAS45347	Chemically pretrea
7	74.8	4.4	9539	24 ABK28180	DNA transcription
8	74.8	4.4	6644	20 AAX33181	Base sequence of t
9	74	4.4	7372	20 AAX33182	Base sequence of t

10	74	4.4	7797	20 AAX33180	Cowpox virus bsr f
11	74	4.4	7996	20 AAX33184	Base sequence of t
12	73.4	4.3	113515	24 ABL34174	Human immune syste
13	72.6	4.3	16033	24 ABL33404	Human immune syste
14	72	4.2	16167	24 ABL70254	Chemically treated
15	72	4.2	16167	24 ABL33083	Human immune syste
16	72	4.2	16167	24 ABL34529	Human metastasis a
17	71.6	4.2	17934	20 AAX02996	Human immune syste
18	71.2	4.2	10620	20 AAX02996	Human IL-1ra BAC c
19	71.2	4.2	14690	20 AAX22303	Human IL-1ra BAC c
20	70.6	4.2	34769	22 AAS46775	Tumour suppressor
21	70	4.1	5750	22 AAS46708	Tumour suppressor
22	70	4.1	5750	24 ABL34008	Human immune syste
23	69.6	4.1	556	24 ABQ36996	Oligonucleotide fo
24	69.6	4.1	556	24 ABQ36997	Oligonucleotide fo
25	68.4	4.0	723	24 ABQ25178	Oligonucleotide fo
26	68.4	4.0	723	24 ABQ25179	Oligonucleotide fo
27	68	4.0	292	24 ABL86605	Human ovarian canc
28	68	4.0	6145	24 ABL32972	Human immune syste
29	68	4.0	16633	24 ABN79984	Human chemically m
30	67.8	4.0	8079	24 ABL92313	Chemically treated
31	67.4	4.0	8404	22 AAS46500	Tumour suppressor
32	67.4	4.0	8404	24 ABL33595	Human immune syste
33	67.2	4.0	556	23 ABV40063	Human prostate exp
34	67.2	4.0	556	23 ABV40163	Human prostate exp
35	67.2	4.0	556	23 ABV42105	Human prostate exp
36	67.2	4.0	556	23 ABV43601	Human prostate exp
37	67	3.9	6668	24 ABL33697	Human immune syste
38	66.8	3.9	174424	24 ABL68122	Ovary cancer relat
39	66.6	3.9	6485	22 AAS46559	Tumour suppressor
40	66.6	3.9	6485	24 ABL33808	Human immune syste
41	66.4	3.9	299	24 ABL86618	Human ovarian canc
42	66.4	3.9	563	23 ABV50724	Human prostate exp
43	66.4	3.9	6338	24 ABL33126	Human immune syste
44	66.4	3.9	16200	24 AAD36438	Mouse L66 genomic
45	66.2	3.9	307	24 ABL86438	Human ovarian canc

ALIGNMENTS

RESULT 1
AAS45477/c
ID AAS45477 standard; DNA; 7657 BP.

AC AAS45477;

DT 18-DEC-2001 (first entry)

Chemically pretreated complementary DNA associated with cell cycle #91.

Cell cycle: human; CpG dinucleotide; cytosine methylation; HIV; aging;
human immunodeficiency virus; neurodegenerative disorder; solid tumour;
graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
PCR primer.

OS Homo sapiens.

XX WO200168911-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02945.

XX 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 30-JUN-2000; 2000DE-1019173.

PR 01-SEP-2000; 2000DE-1043529.

XX (EPIG-) EPIGENOMICS AG.

Db	2568	TATCAGCTATAAAGATATATTAAGTCTTAAACAAAACCTTTACAGCTTTAAAGATCTCAATTTCTC	2509
Qy	145	ATCAAACTCTATCTCTTATATCATCAACGAGCATTTGACGGAAAAATTCAGGTAAAGAGAGAAA	204
Db	2508	ATCTACTATAAACAATAAAAAA	2449
Qy	205	ATAAGAATGAGAGATAGAGAGATTTCTATGGAAAAAGAAAGAGAGACATGTAGGTGAA	264
Db	2448	AAAAA	2389
Qy	265	CAAAATAAAGAGATATCATGATATATTTATGAGAGGTGCTGAAGATTTATTTTAGGAGAG	324
Db	2388	AAAAAAACCAATATAAAAAA	2329
Qy	325	GGAGAGAAATAGAAAAAGAAATGACATGGTGAATCTGAAGAAGATCAATTTGTGTAA	384
Db	2328	ATAAAAAA	2269
Qy	385	AGATGAAGAGAGAAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATGATAAGAAAGAAAC	444
Db	2268	AAAAAATAAAAAA	2209
Qy	445	AAAAAGAGGAAGAAAGAAAGGCTAAAAATAGACTAACTATTTCGCAAAA	496
Db	2208	AAAAAATAAAAAA	2157
RESULT 4			
AAT72882			
ID	AAT72882 standard; cDNA; 19124 BP.		
XX			
AC	AAT72882;		
XX			
DT	12-SEP-1997 (first entry)		
XX			
DE	Plasmodium var-7 gene.		
XX			
KW	DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy;		
KW	Duffy binding like gene; Duffy antigen binding protein; erythrocyte;		
KW	DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;		
KW	Plasmodium; ss.		
XX			
OS	Plasmodium vivax.		
OS	Plasmodium falciparum.		
XX			
FH	Key		
FT	exon		
FT	Location/Qualifiers		
FT	7317..15139		
FT	/*tag= a		
FT	/number= 1		
FT	15140..16205		
FT	/*tag= b		
FT	/number= 1		
FT	exon		
FT	16206..17552		
FT	/*tag= C		
FT	/number= 2		
FT	/note= "no stop codon given"		
XX			
XX	WO9640766-A2.		
XX			
PN	19-DEC-1996.		
XX			
PD	07-JUN-1996; 96WO-US09508.		
PF			
PR	07-JUN-1995; 95US-0487826.		
XX			
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES..		
PA			
XX	Chitnis C. Miller LH, Peterson DS, Sim KL, Su X;		
PI	Wellens TE;		
XX			
XX	WPI; 1997-052231/05.		
DR	P-PSDB; AAW22475.		
XX			

PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 XX Claim 4; Page 56-61; 96pp; English.
 PS
 CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).
 XX
 SQ

Query Match 4.5%; Score 76; DB 18; Length 19124;
 Best Local Similarity 50.0%; Pred. No. 1e-05;
 Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;
 QY 52 ATTGGATGACGACAAATATTATATATACAGTGTAAATCAATCTTCTTCTTAAAT 111
 DB 15440 ATTTTAAATGAAAGAAAGAAATGAAATATAAATAAATTTTATTAATAAATAA 15499
 QY 112 ACATGCTTTTAAATCAGACATTTGTTTAAATCAATCTTCTTCTTATATCAAC 171
 DB 15500 AAAAAAAGAAAGAGAGAAATTTTAAATCAATCTTCTTCTTATATAAATAA 15559
 QY 172 GACATTGACGAAATTCAGGTAAAGAGAAATTAAGAAATGAGAGATGAGAGATTTC 231
 DB 15560 AATTTTGATAGATAAATAATGAAAGAGATTATCAAAAAAATTTAAATAAATAA 15616
 QY 232 TATGGAAGAAAGAGAGACATGTAGTGAAACAAATTAAGAGATATGATATATT 291
 DB 15617 TATATAAATAAATAATGATTATAAATAAATAAATAAATAAATAAATAAATAA 15676
 QY 292 TTATGAGAGGTGGTGAAGATTATTTAGGAGAGCGAGAGAGAAATAGAAAAAGAAATGA 351
 DB 15677 AAAAAAATAATATATATATATAAACAATAAATAAATAAATAAATAAATAAATAA 15736
 QY 352 CATGGTGAATCTGAAGAGATGAATGTTTAAAGATGAAGAGAGAAAGAAATCCCATG 411
 DB 15737 AAATATATATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15796
 QY 412 GCTAAAGTCTCGTAAAGAGATGAAGAAAGAAATTAAGAAAGAGAGAAAGAGGC 471
 DB 15797 AAATAAATAAATTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15856
 QY 472 TAAATAGACTACTA 487
 DB 15857 AAAAAAATAAATAA 15872

RESULT 5
 AAZ98287
 ID AAZ98287 standard; DNA; 19124 BP.
 XX
 AC AAZ98287;
 XX
 DT 13-JUN-2000 (first entry)
 XX

DE Plasmodium var-7 polypeptide encoding DNA.
 XX
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KW DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 KW protozoicide; var-7; ds.
 XX
 OS Plasmodium sp.
 XX
 PN US5993827-A.
 XX
 PD 30-NOV-1999.
 XX
 PF 07-JUN-1995; 95US-0487826.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Sim KL, Chitnis C, Peterson DS, Su X, Wellem's TE, Miller LH;
 XX
 DR WPI: 2000-194198/17.
 DR P-PSDB; AAY77904.
 XX
 XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX
 PS Disclosure; Columns 91-108; 93pp; English.
 XX
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria. The present sequence represents the var-7 polypeptide encoding
 CC DNA.
 XX
 SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 4.5%; Score 76; DB 21; Length 19124;
 Best Local Similarity 50.0%; Pred. No. 1e-05;
 Matches 218; Conservative 0; Mismatches 215; Indels 3; Gaps 1;
 QY 52 ATTGGATGACGACAAATATTATATATACAGTGTAAATCAATCTTCTTCTTAAAT 111
 DB 15440 ATTTTAAATGAAAGAAAGAAATGAAATATAAATAAATTTTATTAATAAATAA 15499
 QY 112 ACATGCTTTTAAATCAGACATTTGTTTAAATCAATCTTCTTCTTATATCAAC 171
 DB 15500 AAAAAAAGAAAGAGAGAAATTTTAAATCAATCTTCTTCTTATATAAATAA 15559
 QY 172 GACATTGACGAAATTCAGGTAAAGAGAAATTAAGAAATGAGAGATGAGAGATTTC 231
 DB 15560 AATTTTGATAGATAAATAATGAAAGAGATTATCAAAAAAATTTAAATAAATAA 15616
 QY 232 TATGGAAGAAAGAGAGACATGTAGTGAAACAAATTAAGAGATATGATATATT 291
 DB 15617 TATATAAATAAATAATGATTATAAATAAATAAATAAATAAATAAATAAATAA 15676
 QY 292 TTATGAGAGGTGGTGAAGATTATTTAGGAGAGCGAGAGAGAAATAGAAAAAGAAATGA 351
 DB 15677 AAAAAAATAATATATATATATAAACAATAAATAAATAAATAAATAAATAAATAA 15736
 QY 352 CATGGTGAATCTGAAGAGATGAATGTTTAAAGATGAAGAGAGAAAGAAATCCCATG 411
 DB 15737 AAATATATATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15796
 QY 412 GCTAAAGTCTCGTAAAGAGATGAAGAAAGAAATTAAGAAAGAGAGAAAGAGGC 471

PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT tumours or cancer
XX
PS Claim 1; SEQ ID No 54; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 other;

Query Match 4.4%; Score 74.8; DB 24; Length 9539;
Best Local Similarity 48.4%; Pred. No. 1.6e-05;
Matches 208; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
QY 67 AAATATTAATATATACAGTGTAAACATGTTTCTTAAATATACATGATTTTAAAA 126
DB 589 AAAAAAAAAATAAACAACAAAAAATAAACAACAAAAAATAAACAACAAAAAATAA 530
QY 127 TCAGACATTTGTTTAAATCAATCTAATCTTATATCAACAGCATTTGACGGAATA 186
DB 529 AAAAAAAAAAATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 470
QY 187 TTCAGTAAAGAGAGAAATAAGAAATCAGAGATAGAGAGATTTCTATGGAAAGAAAG 246
DB 469 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 410
QY 247 AGAGAACATGTAGGTGAACAAATAAAGAGATATGATGATATATTTATGAGAGGTGGT 306
DB 409 AAAAAACGAAAAAATAAACAACAAAAAATAAATAAATAAATAAATAAATAAATAA 350
QY 307 AGATATTTTGGAGAGGGAGAGAGATAGAAAAAGAAATGACATGGTGAATCTGAA 366
DB 349 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 290
QY 367 GAAGATGATTTGTTTAAAGATGAAGAGAGAAAGAACTCCATGGCTTAAAGTCTCTGTA 426
DB 289 AAAAAACACACAAAAAATAAACAACAAAAAATAAATAAATAAATAAATAAATAA 230
QY 427 AGAAGATGAAAAAGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486
DB 229 ATAAACGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 170
QY 487 ATTGCCAAAA 496
DB 169 AAAAAAATAA 160

RESULT 8
AA333181
ID AA333181 standard; DNA; 6644 BP.
XX

AC AAX333181;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRx-ires-bsr.
XX
XX Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
XX inflammatory disease; bsr.
OS Synthetic.
OS Cowpox virus.
XX
XX WO9913073-A2.
PN
XX 18-MAR-1999.
PD
XX 07-SEP-1998; 98WO-JP04010.
PF
XX 08-SEP-1997; 97JP-0259235.
PR
XX (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PA
XX Hamada H;
PI
XX WPI; 1999-243728/20.
DR
XX New apoptosis-resistant virus-sensitive cell
PT
XX
XX Example 1; Page 38-41; 51pp; English.
XX
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the
CC plasmid pRx-ires-bsr, which contains the cowpox virus bsr gene, and
CC is used in an example from the present invention.
XX
SQ Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 other;
Query Match 4.4%; Score 74; DB 20; Length 6644;
Best Local Similarity 48.4%; Pred. No. 2.1e-05;
Matches 206; Conservative 0; Mismatches 220; Indels 0; Gaps 0;
QY 107 AAATATCATGCAATTTTAAATCAGACATGTTGTTTAAATCAATCTAATCTCTTTATC 166
DB 4018 AA 4077
QY 167 ACAACAGCATTTGCGGAAATTCAGGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 226
DB 4078 AA 4137
QY 227 ATTTCTATGAAAAAGAGAGAGACATGCTAGGTGAACAAATAAAGAGATATGATGAT 286
DB 4138 AA 4197
QY 287 ATATTTTATGAGGTGCTGAAGTATTTTAGGAGAGGGAGAGAGAGAGAGAGAGAGAG 346

WPI; 1999-243728/20.

New apoptosis-resistant virus-sensitive cell

Example 1; Page 34-38; 51pp; English.

The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the cowpox virus bsr gene which is used in an example from the present invention.

Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 other;

XX Human immune system associated gene SEQ ID NO: 1056.
 DE Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytotostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PP 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 1056; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 16167 BP; 3894 A; 285 C; 4136 G; 7838 T; 14 other;
 Query Match 4.2%; Score 72; DB 24; Length 16167;
 Best Local Similarity 46.7%; Pred. No. 6.3e-05;
 Matches 207; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
 QY 25 AAATATTATTATTCATGAGTTTATTTGGATAGCATGACAAATATTATATATATCAGT 84
 DB 5977 AATAATATAAAATCTTAATTAATTTATAAAATTTAACTTATTAATAATTTTAAAAAT 5918
 QY 85 GTTAATAACATGTTTCTTCTTAAATACATGATTTTAAATCAGACATTTGTTTAA 144
 DB 5917 TAATATAATTTCAATTTATTAATCTTAACTTTAAATATACCTCATTTTATTTCTACT 5858
 QY 145 ATCAAAATCTAATCTTTATATCAACGACATTCAGCGAAATTCAGGTAATAAGAGAAA 204
 DB 5857 TTTCTCACACCAAAACGNNNAACAAACGAAACAAACAAACAAACAAACAAACAAACAA 5798
 QY 205 ATAAGAATGAGAGATGAGAGATTTCTATGGAAAGAAAGAGAGAGATGTAGGTGAA 264
 DB 5797 AACGAAACAAACAAACAAACAA 5738
 QY 265 CAAATAAGAGATATCATATATTTTATGAGAGGTGCTGAAGATTATTTTAGGAGAG 324
 DB 5737 ACGANAAAAACGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 5678
 QY 325 GGAGAGAGAAATAGAAAAAGAAATGACATGTGTAATCTGAAGAAGATGAATTTGTTAA 384
 DB 5677 AACGCGCAACAAAAA 5618

QY 385 AGATGAAGAGAGAGAGAGAACTCCATGGCTAAAGTCTCGTAAAGAGATCAAAAAAGAAAC 444
 DB 5617 AA 5558
 QY 445 AAAAG 467
 DB 5557 AA 5535

Search completed: November 28, 2002, 23:12:13
 Job time : 805 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 21:46:16 ; Search time 3105 Seconds
(without alignments)
15933.903 Million cell updates/sec

Title: US-09-733-685-3
Perfect score: 1700
Sequence: 1 tgggtttttattgataaca.....tttaacgacttaacgaacatat 1700

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vt:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vt:*
- 30: em_vt:*
- 31: em_vt:*
- 32: em_vt:*
- 33: em_vt:*
- 34: em_vt:*
- 35: em_vt:*
- 36: em_vt:*
- 37: em_vt:*
- 38: em_vt:*
- 39: em_vt:*
- 40: em_vt:*
- 41: em_vt:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	1700	100.0	83650	8	AB023041	AB023041 Arabidops
C 2	576.2	33.9	92620	8	AB026636	AB026636 Arabidops
C 3	540.2	31.8	83646	8	AB005248	AB005248 Arabidops
C 4	534.6	31.4	94487	8	AC012394	AC012394 Arabidops
C 5	534.6	31.4	100806	8	AC015450	AC015450 Arabidops
C 6	499.8	29.4	104386	8	AT032417	AT032417 Arabidops
C 7	499.8	29.4	179771	8	AT032417	AT032417 Arabidops
C 8	497	29.2	95519	8	AF071527	AF071527 Arabidops
C 9	497	29.2	116448	8	AC005142	AC005142 Arabidops
C 10	497	29.2	159629	8	AT032417	AT032417 Arabidops
C 11	369	21.7	369	8	AT032417	AT032417 Arabidops
C 12	330.8	19.5	95190	8	AT032417	AT032417 Arabidops
C 13	103.2	6.1	7218	6	166494	Sequence 14
C 14	90.2	5.3	171763	2	AC115878	AC115878 Mus muscu
C 15	89.2	5.2	280982	2	AC125184	AC125184 Mus muscu
C 16	87.4	5.1	161362	2	AC101939	AC101939 Mus muscu
C 17	87.4	5.1	217779	2	AL844515	AL844515 Mus muscu
C 18	86.8	5.1	202872	2	AC016160	AC016160 Homo sapi
C 19	86.6	5.1	166299	2	AC127244	AC127244 Mus muscu
C 20	85.6	5.0	44735	9	AC092315	AC092315 Homo sapi
C 21	85.6	5.0	48623	2	AC130469	AC130469 Homo sapi
C 22	85	5.0	178038	10	AL672308	AL672308 Mouse DNA
C 23	84.6	5.0	422	5	PD0004	PD0004 P.domesticu
C 24	84.6	5.0	61953	2	AC101966	AC101966 Mus muscu
C 25	84.6	5.0	177648	10	AL732328	AL732328 Mouse DNA
C 26	84.6	5.0	228944	2	AL844530	AL844530 Mus muscu
C 27	84.4	5.0	222556	2	AC126558	AC126558 Mus muscu
C 28	84	4.9	417	5	AF044763	AF044763 Cecropia
C 29	84	4.9	202756	2	AC119804	AC119804 Mus muscu
C 30	84	4.9	218898	2	AC114903	AC114903 Mus muscu
C 31	83.6	4.9	162803	2	AC117818	AC117818 Mus muscu
C 32	83.6	4.9	206538	2	AC102446	AC102446 Mus muscu
C 33	83.4	4.9	239851	2	AC101712	AC101712 Mus muscu
C 34	83.4	4.9	247340	2	AC121591	AC121591 Mus muscu
C 35	83.2	4.9	64789	2	AC083839	AC083839 Homo sapi
C 36	82.8	4.9	167364	10	AC121590	AC121590 Mus muscu
C 37	82.8	4.9	220275	10	AL732404	AL732404 Mouse DNA
C 38	82.6	4.9	48699	2	AC115677	AC115677 Dictyoste
C 39	82.4	4.8	178413	2	AL805954	AL805954 Mus muscu
C 40	82.4	4.8	203295	10	AL663049	AL663049 Mouse DNA
C 41	82.2	4.8	204829	2	AC113318	AC113318 Mus muscu
C 42	82	4.8	163132	2	AC097180	AC097180 Rattus no
C 43	81.6	4.8	93661	9	AC107463	AC107463 Homo sapi
C 44	81.6	4.8	106815	10	AL645809	AL645809 Mouse DNA
C 45	81.6	4.8	120635	2	AC127463	AC127463 Ornithorh

ALIGNMENTS

RESULT 1

AB023041/c

LOCUS

DEFINITION

AB023041 BA000014

VERSION

AB023041.1 GI:4220640

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

clone:MPE11

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (sites)

AUTHORS

Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.

AB023041 83650 bp DNA linear PLN 27-DEC-2000
Arabidopsis thaliana genomic DNA, chromosome 3, PI clone: MPE11.

AB023041 BA000014

AB023041.1 GI:4220640

Arabidopsis thaliana (strain:Columbia) DNA, clone:PIB:Mitsui PI
clone:MPE11.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (sites)

AUTHORS

Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.

TITLE	Structural analysis of Arabidopsis thaliana chromosome 3. I.	CDS	SWRPEFLPAE"
JOURNAL	Sequence features of the regions of 4,504,864 bp covered by sixty		complement(join(10090..10408,10649..10804,10916..10962))
MEDLINE	PI and TAC clones		/note="gene_id:MPE11.3
REFERENCE	DNA Res. 7 (2), 131-135 (2000)		similar to unknown protein
AUTHORS	2 (bases 1 to 83650)		sp P42744"
TITLE	Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.		/codon_start=1
JOURNAL	Submitted (01-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research		/evidence=not_experimental
	Institute, Department of Plant Gene Research; 1532-3, Yana,		/protein_id="BAB01049.1"
	Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,		/db_xref="GI:9279591"
	Tel:81-438-52-3935, Fax:81-438-52-3934)		/translation="MMSPKAKYDQLMYTTOGTLEBASICLLNCGPIGSSNAKNUVLG
COMMENT	Address for correspondence: kaos@kazusa.or.jp		GVGSTITVEGSKVLIDGKQFHRHAIEQKESISEGFRDENTVFQRHQHSVFORQL
	For the latest information on annotation of this clone, please see		BQNR1AGQTVRPMDIARRIWASIGRMWLSADRTTTCYKEARLITDPPGFRVLRLTSL
	http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=Mpe11		VDRLDFERRLCDI"
	Genes with similarity to proteins in the databases are described in		complement(11450..11818)
	'product' or 'note' qualifiers. Genes that have no significant		/note="gene_id:MPE11.4
	protein similarity are described as 'unknown protein'.		unknown protein"
	The software programs used to predict genes include: Grail		/codon_start=1
	(Informatics Group, Oak Ridge National Laboratory,		/evidence=not_experimental
	http://combio.ornl.gov/Grail-1.3/),		/protein_id="BAB01050.1"
	GENSCAN (Chris Burge, MIT, http://CCF-081.mit.edu/GENSCAN.html),		/db_xref="GI:9279592"
	NetGene2 (S.M. Hebsgaard, et al., CSS, Technical University of		/translation="MNSLAKKEERVEEDNGKSDGNRGKPSDEVVTVTVEEVDPPKI
	Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and		LRRVHVATRVAKNGVYAGCELPSPKRRKRSQNLGRNSLDGCVGDGDFEINRVGL
	SplicePredictor (Volker Brendel, Stanford University,		QGLGIDLNCRPEPDSVLSL"
	http://grenlini.2001.iastate.edu/cgi-bin/sp.cgi).		16981..17979
	Genes encoding tRNAs are predicted by tRNAscan-SE		/note="gene_id:MPE11.5"
	(Sean Eddy, Washington University School of Medicine, St. Louis,		/codon_start=1
	http://genome.wustl.edu/eddy/tRNAscan-SE/).		/evidence=not_experimental
	This sequence may not be the entire insert of this clone. It may be		/product="AP2 domain transcription factor-like protein"
	shorter because we remove overlaps between neighboring submissions.		/protein_id="BAB01051.1"
	The 5' clone is K9122 and the 3' clone is MLL14.		/db_xref="GI:9279593"
FEATURES	Location/Qualifiers		/translation="MAEKKRRSSIQTNKPKMKKKPPQNLHLPGLSDELTKMKRLR
source	1..83650		FVNDVATDYSSEEEERORRKYVCEIDLFAQAATAQAESESYCOEENNNGVSK
	/organism="Arabidopsis thaliana"		TK1SACSKVLRKSAFVGVRSSTTSKPVGRQKWKAAEIRHPITKVTWLTGY
	/strain="Columbia"		ETLEQAADAVATKRLDALAAATSAASVLSNESGSMISASGSDLDLKLVDSTLD
	/db_xref="taxon:3702"		QACGSKKASGDFDADLOIPMGCFIDDSFIPNACELDFLLTEENNQMLDYGCD
	/chromosome="3"		DDLIIGLECDGSPDFDYSVEIDLGLIGTTIDKYAFVDHTATTTPTPLNATCP"
	/clone="MPE11"		join(21893..22053,22330..22386,22578..22783,22894..23227,
	/clone_lib="Mitsui Pl"		23311..23382,23473..23613)
	/complement(542..764)		/note="gb AAAF23821.1
exon	/note="CDS is reported in Acc# AP000599		gene_id:MPE11.6"
	contains similarity to CHP-rich zinc finger protein		/codon_start=1
	gene_id:K9122.5"		/evidence=not_experimental
	/number=1		/product="homocysteine S-methyltransferase AthMT-1"
	/evidence=not_experimental		/protein_id="BAB01052.1"
	join(1705..2463,2548..2739,2862..2949,3037..3107,		/db_xref="GI:9279594"
	3190..3273,3410..3514)		/translation="MVEKKSALEDLIKKCGCAVVDGFGFATOLEIHGAANDPLWS
	/note="gb AAD55139.1		AVSLIKNPelikRVHMEYLEAGDIVVTISQATIPGLSRGLSIEESLQKSVEL
	gene_id:MPE11.1"		AVEARDRFEKVKSVGHYNRAIVASISGYGAYLADGSEYSGHYGENVSLDKLDF
	/codon_start=1		HRRLOVLVEAGDPLLAFAETIPNKLQAACVELLEEKVQIPAWICFTSVGDEKAPSG
	/evidence=not_experimental		ESFECLPLNKSNNIYAVGNCAPPOFENLRKFAKLTKRAIVVYPSNGEWDGKA
	/product="dihydrolipoamide S-acetyltransferase"		KOWLPSCGDDREFEATKRDIGAKLIGCCRTTPTSTINATSRDLKRR"
	/protein_id="BAB01047.1"		complement(2777..27552)
	/db_xref="GI:9279589"		/note="gene_id:MPE11.7
	/codon_start=1		unknown protein"
	/evidence=not_experimental		/codon_start=1
	/product="dihydrolipoamide S-acetyltransferase"		/evidence=not_experimental
	/protein_id="BAB01047.1"		/protein_id="BAB01053.1"
	/db_xref="GI:9279589"		/db_xref="GI:9279595"
	/translation="MTVRSKIRIEFMPALSTMEGRKIVSWIKTEGEKLAKGESVVVV		/translation="MTHAREWRSLLTTLMLVLLSYMLHLCVYKVRGAIREFETP
	ESDKADMEVFYDGYLAAIVVGEGETAPVGAALGLIAETEAEIEPAKSKAASKSSSS		ASGKRQEDLMKKYFGAGKFPDVSFVGKGISSEKRVISCPDPLN"
	VAEAVPVPPTSPAPAIQAOPAVTAVSDGPRKTATPYAKKLAKQHKVDIESVAG		30765..31883
	TGPFGRITASDVEAAGTAPSKSIAPPPPPPPVAKATTNLPPLPDSIVPFTA		/note="gb AAFI6548.1
	MQSAVKNNIESLSPVTFRCVYPVNTDALLALEYKPKGYVTALLAKAGMALAQH		gene_id:MPE11.8
	PVYNASCKDGKSYNSINIAVAINGLITPVLQADKLIDLILLSQKRWELVGKA		similar to unknown protein"
	RSKQLOPHEYNSGFTLSNLGMFGVDRFDAILPPGQAIMVAGSKPTVVVADKDGFFS		/codon_start=1
	VKNMLVNTADHRIVGADLAALFQTFAKIENPDLSLT"		/evidence=not_experimental
	complement(4594..5106)		/protein_id="BAB01054.1"
	/note="gene_id:MPE11.2		/db_xref="GI:9279596"
	unknown protein"		/translation="MPKERRSVSLDKYKRSPLCCEASLAKPSQKQKWEARCP
	/codon_start=1		NCMPHPINGILLICSSYNGCRPYMCDTSHRNSCDFQFRKSKPSLSLREBEES
	/evidence=not_experimental		VEPTMEDVSDSTAVNLGAAESTIVVLDSDGERGEVEEVEEVEVEEVEEIV
	/protein_id="BAB01048.1"		TTEODQKNPKOKLTCPLCRGHIKEWVVKAAFCMNSKHSKSCCTCFSGYSOLR
	/db_xref="GI:9279590"		KHARLLHGVPRSEADPERQSWRRLERQSDIGLISLTLQSSFGGDEISDNDGFLFAD
	/translation="MEDLLEERLRTDVSNGKRVKRGDLSDVFKRLRDLDFDSDGL		
	DPVSDQLDSVMKSFENELSTLTAAALSGSETQPOLGYLFEASDDELGLPPLTPQTL		
	PPSCBETVTELVRASDSEVGLCGFDHVTETFGPCDLDGDLFEYFDGCLDSDGLE		


```

Qy 1501 GSCCTTTCCCTGGCCCGCTTTAGTAATATTCAGAGTAGGATAGGCTATGGCCCTGATAGGGAGT 1560
|||||
Db 11180 GSCCTTTCCCTGGCCCGCTTTAGTAATATTCAGAGTAGGATAGGCTATAGGGAGT 11121
|||||
Qy 1561 CCAAAAAGAGCGGCGCATTCGGGTTGGGTGGCTTTGGAACTTTGGATTGTGGATTAGT 1620
|||||
Db 11120 CCAAAAAGAGCGGCGCATTCGGGTTGGGTGGCTTTGGAACTTTGGATTGTGGATTAGT 11061
|||||
Qy 1621 CATGTTTATCTATTAAATGCTCGGGACTTCGTGACGACGCGCTGTGTTCTTCTTCCCTCTG 1680
|||||
Db 11060 CATGTTTATCTATTAAATGCTCGGGACTTCGTGACGACGCGCTGTGTTCTTCTTCCCTCTG 11001
|||||
Qy 1681 TTTAGGACTTACGACATAT 1700
|||||
Db 11000 TTTAGGACTTACGACATAT 10981
|||||

RESULT 2
AB026636 92620 bp DNA linear PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K14A17.
ACCESSION AB026636 BA000014
VERSION AB026636.1 GI:4757392
KEYWORDS Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC
SOURCE clone:K14A17.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
p1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
20277480
2 (bases 1 to 92620)
Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.
Direct Submission
Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=K14A17
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Graal-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremlin1.zool.kastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MH15 and the 3' clone is MCE21.
Location/Qualifiers
1..92620
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="K14A17"
/clone_lib="Mitsui TAC"
complement(201..2180)

CDS

```

```

/note="gene_id:K14A17.1"
/codon_start=1
/evidence=not_experimental
/product="salt-inducible protein-like"
/protein_id="BAA94973.1"
/db_xref="GI:7670019"
/translation="MRGPASSASRIATAAFAAKSLNAS"SYNPKLSKTLNSSGKPTNP
LNQRYISOVIERKDWFLINQEFTHRIGLNTREVISV QNQNPLHSIRFYLWVSNP
APYKADQSLKSLGNALFRKGPLLSMELLKLNDS:RTSBLKCVLIGSWRIGL
KVCNDVPAQJSELGMPSTRLYNALIDALVKSNSLDLVLKRNHMSDCKPDRFTY
NILLHGCKGVDEARLVKQMEQEGNRPNI IYTLIDGFLAGRVDEALKOLEMM
RVKLNPEATIRTFVHGIFRCLPCKAFVLVGMER NLRQVSGVDAYLVLCUSNS
MAKETQPLRKIGERYIPDSFTNAAMSCLLKGDLV,TCRIFDSVRGKPGFNG
YLVLVALLNAQRFSEGRDYLKQMGVGLLSYSYNAVINCLCKARRIENAAFLFE
MODRGISPLNLTENTFLSGYSVRGDKVKGVLKLLV GKFPDVTFFSLINACRA
KEIKDAPCFKEMLEWGLEPNEITYNLIIRSCSTGDI RSVKLFKAKENGSLSPDLY
AYNATIOQFCMKRVKKAELLKTLRIGLCKFCNFIYS LIKALSSEGRESEAREMS
SIERHGCVDSYTRKLVBEELDLKRSGLSRSTVSA"
Join(3906..4124,4238..4462,4536..4649)
/note="embI: CAB82745.1"
gene_id:K14A17.2
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA94974.1"
/db_xref="GI:7670020"
/translation="MAKQDPTSSNTMLPLVGEI EVRPQJLDTTGTVDKATGNKVF
QKVPFLGLHNKRILVDPNDSPITVMKMKVTSFHRNCA VRGSLDDDKIPTVRKSSIV
OLKTRVEFLKHNQTRRESSCDFTLKGRMKRAITTYVALSTKITAQVYEGHERLVATI
YPNVDYAIRVLIIFLDLINMGTCI"
complement(join(5031..5929,6011..5601,6984..7211))
/note="gene_id:K14A17.3"
/codon_start=1
/evidence=not_experimental
/product="AMP-binding protein"
/protein_id="BAA94975.1"
/db_xref="GI:7670021"
/translation="MAATPKRIDDDPKIPANITALTPLNFDRAAVVHPTKRSVING
SREYTRQYDRCRLASALADRSIGSGSIVFVALTANIPAMYEAFHGVPMCGAVIN
CVNTRLNAPTVAFLHSQSSVIMVDQBFITLHETLMEERAGSFPKRLPILVIGD
HTCAPELSNRALSKGAIYEDFLATGDPNI PWCPHAFI FQSTIALGYSQTTSKPGVV
LHKGAVTALSNSPLIMGMOGAVLATLMEHCNLCV POSLAVLSGTSICLRGVV
KEYVSMIAKYKTHFCAPAVYLVNAILVNAKEL IILIPVHVWVTGAAPPPSVLFSA
NOKGRFARVATGLSETGYTPSVCAWKPEWDSI:EL:CA:LNARQVYTRTGEQDQVID
TGTGKVPADGKTAGEIVFRGNMVNKGILKNPI AN:ET AGGFHSSDIAVHKPDNYI
ETIKRSKDVIIISGGENISSEVENVYVHPAVI EA:VAVRDERWQSRSCAFVFLKSD
YEKHONQLAQD IMKFCREKLPAYVWPKSVVFLPKRLATGKIQKILHATKAKEMGPV
PRSLI"
complement(join(8463..8892,8977..9136,9230..9641))
/note="gene_id:K14A17.4"
/codon_start=1
/evidence=not_experimental
/product="basic chitinase"
/protein_id="BAA94976.1"
/db_xref="GI:7670022"
/translation="MVSPLFSLLLTVALVVIQTGLVNAEDSEPSSTRKPLVKIV
KKGKLCQWCKGKMSVCCNHTISDFETQYFENLFSKRNPSVAHVAGFWDYSFIF
AAAEYQPLGFGTAGELKQKMEVAFAFGLHGVSGTSCGYVATGPGFLAWGLKEMSP
DQLYCDDYYKLTYPCTPGVSYHGGALPVYWNYGOTGEALKVDDLHSHPEYLENNAT
LAFQALIRWMTTPPKHLPSADHVEVGKWKPTKNDTAARTPGFGATINVLGQICN
SGFDNEMNNIVSHLYILDLIGVREAGPHEKLSAQDQPPSSSSAPPSSSSSS"
Join(13759..13822,14371..14445,14592..14707,14800..14888,
15014..15059,15238..15433,15520..15845,15928..17178,
17360..17713)
/note="gene_id:K14A17.5"
/codon_start=1
/evidence=not_experimental
/product="transcription factor-like protein"
/protein_id="BAA94977.1"
/db_xref="GI:7670023"
/translation="MDGGLGLRIGTSETHGFTLQDLVDOTMLEAKSRWRLPNEIHA
ILCGRIILDFRKMILNFRKGDHNNKKKDKGRTVKEAHEHLKVGNIEERHVIYAGEDN
TTFVRRCYWLILDKARENILVHVYRDTQEAATISGDSISSPISVSISVFPNVAADID

```


TVVRNHDLSLHDLINTLWDELVTPTDLNNOSAPTVDNLSYFTEPLQNAANGTAHGNH
 TVADGSLDALLNDGQPSRESFGRMNSFISESNGSLDSDPSFEPWMPRODPLAPQAVF
 HSHSNIFQVFNITDPSPAWYSESKTKILVTGLDHSYOHLSRLNYCVGDFCVPA
 EYLAQGYVRCIIPPHSPGMVNLXSADGHKPISQCFEFHRAVPLDKTVPEDNQDSK
 WEEFQVRLSHLFTSSNKLSSLSKISPHNLDAKTLKASKTNHLLNSWALYLVKSIQ
 GKNYSFOADHLFLFSLSKNRLKLEMEKVLGRNLDYDYSKGLGVHLCASIGYTWQ
 VOLFSLSGLNFRDQGVWALHAAVYGREKNVAALISAGARPNIATVSTKDNLGGC
 MAADLAQONGYDGLAAVLAEPKLAQFQPMKIAQNTIGDLACKAEMLNQGLTPEDEQ
 SLKDALAAYTAEEAARIQAFERAKAARSSVIOFANKKEEAKSIIAAMKIQAF
 RKVDTRKRRIKAAIKICQFQTKWIKREYLNMRQARIQAARFGLQARQYKKILWSV
 FRSKKAQDYRRMKLTHEEAQVNLFTLNLSPFGKKNRR"
 complement(join(18170..18172,18407..18490,18736..18807,
 18977..19090,19168..19281,19387..19487,19590..19668,
 19755..19805,19906..19989,20365..20431,20539..20669,
 20916..21040,21126..21231,21349..21483,21624..21770,
 21860..22171))
 /note="gene_id:K14A17.6"
 /pseudo
 /codon_start=1
 /evidence=not_experimental
 /product="liponamide dehydrogenase"
 join(29096..29206,29695..29826,29925..30009,30149..30312,
 30830..31267)
 /note="contains similarity to similar to ubiquitin
 conjugating enzyme"
 gene_id:K14A17.7"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAA94978.1"
 /db_xref="GI:7670024"
 /translation="MADERYNRKNPAVKRIQEVKEMOANSDDFMSLPLEINIFEWQ
 FAIRGPDTEFEGYVHGRIQPADYPPKPSFMLLTPNGRETNKICLSISNVHPE
 HMOPSVNLTVALIAFMPTSPNGALGSDVDPKDRRLAKSKRETPKYSPQRQ
 LIDEIHQYILSKATVYKPLPLECSOAPSIVSEANSOPEQAIIVTRESATTDTI
 VDDQIIEETAEAVNTAASVPAAPLPAVEYVSKASVSGEQEMARRAAQKPPVDDRLFT
 WAAGYITAIMLLLLKKFKISNGYSTGFEMDDQS"
 join(32260..32364,33121..33412,33506..33750,33837..34103)
 /note="gb|AAD26962.1
 gene_id:K14A17.8
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAA94979.1"
 /db_xref="GI:7670025"
 /translation="MGKSSNIVFDRIKBRQMLSPFKIFORADLSSDCMRALPYSFIS
 KYSEKDFSVKMWIQAQWKTQWVSKNPTVYIETRGWDLVSDNALQNEFIFFTH
 RGNMVFHNYEQNGLEMRKPKFQTMGFPSSGIKKEGNSLIDVKKKEEDESQRA
 EFLVRRKKTEDSKSKMTRNKRKSKSKQVLDGVPEFKITIRKSYLKFLAIPK
 HFVDDHIPNKSIFTIRHPNGGSKWVLCVLREITIFSGGYSKLAREFLVMGDKCTF
 KLKPFPEFVLLTSKKNREKMDQCMID"
 complement(join(34528..34725,35213..35322,35680..35863))
 /note="gb|AAF26101.1
 gene_id:K14A17.9
 similar to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAA94980.1"
 /db_xref="GI:7670026"
 /translation="NAESGREGIVAVDFDCSKKALSWAINDVVRDGHLLITIAH
 MNLYEGEQMLHETVGSFPFNPSEFSDAAVMKKYALKPDAETLDLVNTAARKRTIVV
 KLIYWDPRKICAAEQIPLUSSLVNRRGLGKRMIMGVSNNHVNNAVACPVTYVK
 AH"
 complement(join(36309..36416,36556..36846,37250..37324,
 37500..37634,38020..38145,38221..38374,38433..38573,
 38878..38987,39134..39281,39474..39751,39993..40262))
 /note="gene_id:K14A17.10
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAA94981.1"
 /db_xref="GI:7670027"
 /translation="MGDTNGASLIEIGQEEVEPFLAFLDYARTVISPDEDEKEE

Query Match 33.9%; Score 576.2; DB 8; Length 92x20;
 Best Local Similarity 99.5%; Pred. No. 1.1e-90;
 Matches 578; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGGTTTTTATGGATAACATGACAAATATTTATTTATTTTCATGAGTTTATTTGATAG 60
 DB 71476 TGGGTTTTTATGGATAACATGACAAATATTTATTTATTTTCATGAGTTTATTTGATAG 71535

QY 61 CATGACAAATATTAATATATCATGTTTAATACATGTTTGTCTTAPAAATACATGCATT 120
 DB 71536 CATGACAAATATTAATATATCATGTTTAATACATGTTTGTCTTAPAAATACATGCATT 71595

QY 121 TTAATATCAGACATTTGTTTTTAAATCAATCTAATCTCTTATATCAACACGACATTGAC 180
 DB 71596 TTAATATCAGACATTTGTTTTTAAATCAATCTAATCTCTTATATCAACACGACATTGAC 71655

QY 181 GGAAATTCAGGTAAAGAGAAATTAAGATGAGATAGATAGATTCTATGGAAA 240
 DB 71656 GGAAATTCAGGTAAAGAGAAATTAAGATGAGATAGATAGATTCTATGGAAA 71715

QY 241 AGAAGAGAGAACATGTAGGTGAACAAAAATAAGAGATATGATATATTTTATGAGAG 300
 DB 71716 AGAAGAGAGAACATGTAGGTGAACAAAAATAAGAGATATGATATATTTTATGAGAG 71775

QY 301 GTGGTGAAGATTATTTTAGGAGGAGGAGAGAAATAGAAAAAATAAGATGACATGGTGA 360
 DB 71776 GTGGTGAAGATTATTTTAGGAGGAGGAGAGAAATAGAAAAAATAAGATGACATGGTGA 71835

QY 361 TCTGAAGAAGATGAATTTGTTTAAAGATCAAGAGAGAAAGAACTCCATGGCTAAAGTC 420
 DB 71836 TCTGAAGAAGATGAATTTGTTTAAAGATCAAGAGAGAAAGAACTCCATGGCTAAAGTC 71895

QY 421 TCGTAAAGAGATGAAAAAGAAACAAAAAGAGAAAGAAAGAAAGAGAAAGAGAAAGAGAA 480
 DB 71896 TCGTAAAGAGATGAAAAAGAAACAAAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAA 71955

QY 481 CTAACATTTGCCAAATTTCTCTAGCCGACAAATCTATTGCTCCAGGTTATTTGTG 540
 DB 71956 CTAACATTTGCCAAATTTCTCTAGCCGACAAATCTATTGCTCCAGGTTATTTGTG 72015

QY 541 TATCTTTTGAAGTCAAAAGTTATTTTACATATCTCTA 581
 DB 72016 TATCTTTTGAAGTCAAAAGTTATTTTACATATCTCTA 72056

RESULT 3
 AB005248/c
 LOCUS
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MX110.
 ACCESSION AB005248 BA000015
 VERSION AB005248.1 GI:2264320
 KEYWORDS
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1 clone:MX110.

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (sites)
 Sato, S., Kotani, H., Nakamura, Y., Kaneko, T., Asamizu, S., Fukami, M.,
 Miyajima, N. and Tabata, S.
 Structural analysis of Arabidopsis thaliana chromosome 5. I.
 Sequence features of the 1.6 Mb regions covered by twenty
 physically assigned P1 clones
 DNA Res. 4 (3), 215-230 (1997)
 97471969
 REFERENCE 2 (bases 1 to 83646)
 Nakamura, Y.
 Direct Submission
 Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research, 1-32-3, Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail:y.nakamura@kazusa.or.jp,

Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/egcd_graph.cgi?c=MX110
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,

http://compbio.ornl.gov/Grail-1.3/),

GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and

SplicePredictor (Volker Brendel, Stanford University,

http://gremlini.zool.tastate.edu/cgi-bin/sp.cgi).

Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis,

http://genome.wustl.edu/eddy/tRNAscan-SE/).

This sequence may not be the entire insert of this clone. It may be

shorter because we remove overlaps between neighboring submissions.

The 5' clone is MS17 and the 3' clone is MB18.

Location/Qualifiers

1. 83646

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="5"

/clone="MX110"

/clone_lib="Mitsui P1"

join(1791..2912,3016..3318,3584..4660)

/note="gene_id:MX110.1"

/codon_start=1

/evidence=not_experimental

/product="disease resistance protein-like"

/protein_id="BAB09346.1"

/db_xref="GI:9758812"

/translation="MTEKIATDVSNMLNNTSPSRDFDLGMDAHKEMESLCLDS

EVRMIGIGTQSGIKKTIVARLYSOFSENFELSLFMGNIKELMYTRPVCSDEYSAKIQ

LQKQFLQINLHKMDHLGVADRLNDKKVLIVLSDIQSLQIDAIAKATRFWFGHG

SRITITQDQKLLKAHGINIHLYKEFFSAYEAQMCYAFQGNFNDGEELAEVIT

KLHLGLPGRVMGSHRGRMSHEWNAIPRLKILRLDASQSLIKSYDALCEEDKDL

FLHTACFLNQMEVEYDIALSFLDVRQGHLLAEKSLNLKFLSTNCTRIEMHLL

VQLGKDIVRHKPGHQSCIEPGKQFLDARDICEVLTDNGRNWVGIFVLEVNLSQ

LNISERAFDGNLGNKIFRHPDYDESDKLYLPQGLNLPQKRLLEWSPFMTCLPS

NFCYKLYVEIPMKNSKLNQWQNLGNLKMWDLSKHLKELPDLSTNLLEYLIM

SGCLSPVLPSSICKRLKMLSLRGSKLEALPTNLNLSLDYLDLTCCLLKKEPE

ISTIKDLTKTKTAIEVPSTIKSWHLKLEMTSENKLELPHALDITITLKINDTE

MOETPQWKYKISHLTQTLGEGCKRLVTIPQSLDSQLVVTNCESLERLNFSPNHPE

RFLWFLNCFKLNNEARFQIOTSSHTALPSPREVPAFTYRANGSSIMVNI.NHRPLSTT

LRFAKVLVKKIDNKEEAADRTTIPRIENDKIGVDVPWRYRPHVPTILEDHLL

TFQLEADVTSKLLFSFDKDNQAVIKGVQLQ"

complement(join(5504..5620,5688..5736,5836..5905,

5990..6122,6206..6250,6374..6688))

/note="gene_id:MX110.2

pir||T03994

similar to unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB09347.1"

/db_xref="GI:9758813"

/translation="MMASISLPSFCLSYFKMPKTAALTNSPTTIPYNLKKQTRLLH

KLPGLKMEVTEQKRSKERENPLVFGVSHAAWCAENWLFSSGFSDFYSVSL

LQGGSDPLGTAGTQTHASDIADIESNLGSPPLVGHSGFLIVGYLANIVN

KRSIGTENAFPELSGAVMVCVPPSGNSGLVRLYFSKPAARFKVTLAAGKPKSI

PLCRTEFFSQAMDQLVKRYNLAR"

join(7788..7835,8043..8777,9032..9157,9205..9901,

10085..12958)

/note="gene_id:MX110.3"

/pseudo

/codon_start=1

/evidence=not_experimental

/product="non-LTR retroelement reverse transcriptase-like

protein"

CDS

complement(13525..13827)

/note="gene_id:MX110.4

pir||T02618

similar to unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB09348.1"

/db_xref="GI:9758814"

/translation="MKDVLIVYLSIKMTKLRLCELDJNSSETYFOLDKHLKNEHDG

LVLMDVPSQWRCOEOMKRRHVEYVDLMSATEPHIAEVVHQPLFIIDLIIIFRLCS"

join(18792..18841,19220..19412,19494..19579,19875..19959,

20464..20607,20697..20783,20894..2106,21097..21162,

21216..21308,21350..21509)

/note="gene_id:MX110.5

unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB09349.1"

/db_xref="GI:9758815"

/translation="MMSPKQISDDRGSGSHRHTPPIILIGNFFRIWVSVMRYLN

QTAGPVLFLFCCLVPSIIIFIQKPKWGRAIENJOIVPSLINGVITALEYFLLWKG

LKSCGPLRAILSEYSGAVLISGLVLYGRGRHWKALPDKDSTETKEEVEEQALGK

MGMTPEAGILSALRVYIARRVSLKNCQKKRIALIT SATCFELFPVAMMDLIGSS

SKTSELFPFAWAFSLTIIFGILLIFIVYDNIAIENLHMVFSSPRHLMVAGVCIIVMDI

RWNIRGNISRTQEQLRFLVIRRSIKRNPSPFIYVNDVLLNVLVDVNLGSGFIR

FSIFVPSFFFCFLGEAYANLV"

join(26073..30268,30416..31552)

/note="gene_id:MX110.6"

/pseudo

/codon_start=1

/evidence=not_experimental

/product="retroelement polypolyprotein-like"

32202..33620

/note="contains similarity to retroelement polypolyprotein

gene_id:MX110.7"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB09350.1"

/db_xref="GI:9758816"

/translation="NVEGTRNPKRSKTMSTETPSEVAHWTKEDKALKVISPPLP

SPPTFRANREILKMLLETIDERTKMGALITMDH TFEPTVDFFSTMEVTFN

KPIPLAKGINKFKYKAFSISIDPLCEAYCIENKSEISPKFKGIHFWGELASGK

FSCNASISVRHPVIRYALKLHALFGRGEISLIVEMCFEQGYKELLVEDADEN

GGRLVHDDGNDVYGCIFASYLKYRDLNKLIGKIKSYIGGITLPLFVRAGVDLSP

FNALPKREYIDYENLVRSCLSDDAFIYLYTD:EDNLECLIPSKETINVESHEDI

MFLPSKEYQVPLEQFMETEDDLPSPRODFENKGLPEFTVPKTKMEKFLQSV

IKSQRKINKKSVATGKLLKVKKLQPPDYVSETEI VVPSNEEDRDNILESSENNE

TPATSSHSRRNARSADHSMP"

join(36310..37248,37409..37549,37533..37824,38363..38519)

/note="emb|CAB86928.1

gene_id:MX110.8

similar to unknown protein"

/pseudo

/codon_start=1

/evidence=not_experimental

join(39564..40460,40627..40767,40853..41185)

/note="gb|AAD55600.1

gene_id:MX110.9

strong similarity to unknown protein"

/codon_start=1

/evidence=not_experimental

/protein_id="BAB09351.1"

/db_xref="GI:9758817"

/translation="MDLLSNLPYELCHLSLFTTKEALTSVLKSRWRNLAFVFNV

LIVDHDIRELFMDVDRVLALQNSPINKFSIDCSGVDSDRVDCQIONVVRGVSSELN

DSIFDVSVPEDDYNPFPKFNKLVKLGSLVLSWLDQSFILPMLKTLVLSVLSVE

KEFILLHALPALEELVMNNIYKREDVDLIIVSSSEKLTLLTNFTVCTHTLSPDTPIL

AYLUSGYVDDIYAEAKMENTLFEARISLLVVEGDISRVALLINLLEDVYDLQFE

NWYKLMNGIRNRCILYSLEVLSCCESPVFNKLSLSTIKSAENRGWQWAPVLL

RNCPLHETLVLEGLLHVVDKDCDCVSRDYGKRSFSGPVKLEIKGFGQTKEM

NMTKHFLEYFPCCLKELKMYMEENDPAQLRVPEVITGEMMERYKKSSCNVQLLVGKL

YKWTPT"

COMMENT

FEATURES
source

CDS

CDS

CDS


```

tRNAScan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAScan-SE/).
Simple repeats are identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
    source                1..94487
                        /organism="Arabidopsis thaliana"
                        /cultivar="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="1"
                        /clone="F15M4"
                        /complement(967..1017)
                        /rpt_family="AT_rich"
                        /complement(1289..1338)
                        /rpt_family="AT_rich"
                        /complement(4053..4080)
                        /rpt_family="AT_rich"
                        /complement(6943..9510)
                        /gene="F15M4.1"
                        /note="identical to hydroxy methylglutaryl CoA reductase
                        (AA 1-592) GB:X15032 [Arabidopsis thaliana]"
                        /complement(join(<6943..7515,7607..7953,8115..8296,
                        8522..>9510))
                        /gene="F15M4.1"
                        /complement(join(7255..7515,7607..7953,8115..8296,
                        8522..9510))
                        /gene="F15M4.1"
                        /codon_start=1
                        /product="hydroxy methylglutaryl CoA reductase (AA 1-592);
                        9510-7255"
                        /protein_id="AAF16652.1"
                        /db_xref="GI:6554470"
                        /translation="MDLRPPPPVPTNNNSGSRSPQRTSDDHRRRTATTIAPP
                        PKASDALPLPLNTNAVFLEFVSVAAYLLHRWRDKRYNTPLHVVTITELGAIIALI
                        ASYLLGFGDFEVSQFISRSGSDADLDATDDDDHRLVCSPTPIVSVAKLPNP
                        RPPPLPVAVIPNMPDLPTLPMELCRVLEVLHLEIRISNHFHTTITKELSLIMVTAY
                        GPVMSQSGLYSPYMQQYLOVYVPGAVNVPYQYQOLSQITPNHGHTAVQGYG
                        PLGDFYESILGQCCMPGYIQIPVGIAGPLLDGYESVPMATTEGCLVASTNRGC
                        KAMFISGATSTVLDGMPTRAPVVRFAASRASELKFLEFPENFTFLAVFNRRSRF
                        ARLOQKTAAGNVAVYRCCSTGDAMGNMVKQNVLEYLTDDFPDMVDVIGISGN
                        FCSQKPAANWTEGRKSVVCEAVIRGEIVNKLKTSVAALVELANLKNLAGSAVAG
                        SLGFNAHSNVSVAFIANGOPQAVNESQCITMMEALNDGDKDITHSVTPMSIBVG
                        TVGGTQLASQSCDLNLGVKGASTSPGNNRRLATIVAGAVLAGELSLMSIAIAGQ
                        LVRSHMKYSSRDSIGATTTTTTT"
                        /complement(7258..7283)
                        /rpt_family="(CAA)n"
                        /complement(11154..11209)
                        /rpt_family="AT_rich"
                        /complement(11551..11580)
                        /rpt_family="AT_rich"
                        /complement(11933..12004)
                        /rpt_family="AT_rich"
                        12390..13268
                        /gene="F15M4.2"
                        /note="similar to unknown protein GB:AAD30603 [Arabidopsis
                        thaliana]"
                        join(<12390..12391,12491..12629,12730..12861,12929..13019,
                        13099..>13268)
                        /gene="F15M4.2"
                        join(12390..12391,12491..12629,12730..12861,12929..13019,
                        13099..13268)
                        /gene="F15M4.2"
                        /codon_start=1
                        /product="unknown protein; 12390-13268"
                        /protein_id="AAF16653.1"
                        /db_xref="GI:6554471"
                        /translation="MFVDKALYCKELRWETLPLOGSRVYVKLQGLKSNWSYEVKISY
                        PASIPARVSLQLNKHEWELKLNOMRRLLNTEKLIFKAESLKEVNNKAGLNVLVLEP
                        EGIIVAPNSRSHRYFIINIVCEQGLGIPYSSVLLVALCLVSVFIVPRSLPSSL
                        TKDGLRSRSHRSYKGS"
                        14056..15506
                        /gene="F15M4.3"
                        /note="similar to cinnamoyl-CoA reductase GB:CAA56103
                        [Eucalyptus gunnii]"
                        join(<14056..14170,14277..14431,14516..14701,14819..14990,

```

```

15096..15285,15362..>15506)
/gene="F15M4.3"
join(14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..15506)
/gene="F15M4.3"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 14056-15506"
/protein_id="AAF16654.1"
/db_xref="GI:6554472"
/translation="MAYKQKVCVTGAGGTASWVKFLSRGYTVHGTVRDPCDEKND
HLRLDNASNNKLKADLDFDEGLFSAIDGCSVFLASPVFEGELIKPALTGPK
NVLEACTETKVQKVYVSSAAVYNPKVQVAKEDCSWDQYILHSLGKYWRYYL
AKLTREALEWSKRNFAFVYVTCPSVIGPRLQSLILNSGLAGLKFIKGGIKSLSD
ELYLVDVADVADALLVYENREATGYICNSFSYTLISLMEKLNKMPYKRNPFESFTE
VKEKEVRPLSAEKLKNLGKFRPLEETIDDSVVVSFEAAGDLPA"
16281..16361
/rpt_family="(CAT)n"
16955..18423
/gene="F15M4.4"
/note="contains Pfam profile: PF00C76 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
join(<16955..17161,17385..17529,17633..17755,17844..17884,
17978..18215,18332..>18423)
/gene="F15M4.4"
/gene="F15M4.4"
join(16955..17161,17385..17529,17633..17755,17844..17884,
17978..18215,18332..18423)
/gene="F15M4.4"
/codon_start=1
/product="putative RNA-binding protein; 16955-18423"
/protein_id="AAF16655.1"
/db_xref="GI:6554473"
/translation="MAYQVPVPGSGHYNSPFGDITFKVFPVGLAMETQSETLRQHF
EQYGEILPAVVIADKNTGSRGSGYGFVFRDPEAARACADPTIIDGRRANCNLASLG
RPPPLPVAVIPNMPDLPTLPMELCRVLEVLHLEIRISNHFHTTITKELSLIMVTAY
GPVMSQSGLYSPYMQQYLOVYVPGAVNVPYQYQOLSQITPNHGHTAVQGYG
VPGSHILGQGTVSTWTSSMPALQAPYPSGIEPEKAVQSHIIVHSPQFMQSTASDQ
TTR"
complement(19352..20843)
/gene="F15M4.5"
complement(join(<19352..19378,19453..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))
/gene="F15M4.5"
complement(join(19352..19378,19453..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))
/gene="F15M4.5"
/codon_start=1
/product="unknown protein; 20843-19352"
/protein_id="AAF16656.1"
/db_xref="GI:6554474"
/translation="MAAISPWLSSPQSFNSPRVITDSRRCSISAAISVLDSNNEQ
HRISRDHVGRKRDVLMQIASVVEFLPLAISPAFAETNASEAFVYVTDENKFEISI
POVSTAITGLGDFTRMESFGKVEAFETLVSLGDRSWKPVGVTKLIDRSRSGEY
VLEYTLQNPGEARKHLYSAIGMATNGWYNRLYTVTQPTDESAEQSKIOKTVKSF
FI"
complement(20998..21026)
/rpt_family="AT_rich"
complement(21140..21167)
/rpt_family="AT_rich"
22462..23090
/gene="F15M4.6"

Query Match      31.4%; Score 534.6; DB 8; Length 94487;
Best Local Similarity 96.5%; Pred. No. 1.8e-83;
Matches 557; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

QY 1 TGGGTTTTATTGGAACATGACAAATATTATTATTTCATGACTTTTATTGATAG 60
|||||
Db 1917 TGGGTTTTATTGGAACATGACAAATATTATTATTTCATGACTTTTATTGATAG 1976
|||||

QY 61 CATGACAAATATTAAATATCATCTGTTAATAACATGTTTGTTCCTTAAATACATCAATT 120
|||||

```



```

/gene="F1466.3"
/codon_start=1
/product="putative phosphate transporter; 18176-15417"
/protein_id="AAG51941.1"
/db_xref="GI:12323970"
/translation="MPGLSLSDAARIQWYHFKAIIVAGMGLFTDYLDFCIAPIM
KMSIQIYKDSITALLSTSYAIALIGTALGOLIFGYLGDVRGRKRVGLSLIMVF
SSRCGGSVCTVRSQVMVSLGPFVILGIGDYPLSATINSEFANKRTRGAFIAA
VFMQGLIIMSSAVTHVVCIAFKNAGESSEKTNVAGLETIAPPESDIARLLILIG
ALPAALFFYRMLMPETARYTALVENNVVAAQDMQRMVSVMISQITDESSELEQV
PSSSYKLFRRRLSLGRDLFAASANFLVDVFTSNLLLSQIFNFNKNPLNSTNV
YDSAFYAKIAAIVAACTIPGYWFTYIFIDKIGRKIQMMGFLLMAVLYLVAGIPYS
WYWKHKTNKNGVLYGLIFFNSFGPNTTFTIIPAEFPFRFSTCHGISGAAGKF
GAIVGTGFWLWATHREEDGFPDVKRVRIAFILILGVCYIAGMIVTVLFTRETMGRSLE
ENEDIYSTSGSSPANELLARROY"
/complement(18673..19301)
/gene="F1466.4"
/note="predicted by genscan+"
/complement(join(<18673..18926,19124..>19301))
/gene="F1466.4"
/complement(join(18673..18926,19124..19301))
/gene="F1466.4"
/codon_start=1
/product="hypothetical protein; 19301-18673"
/protein_id="AAG51943.1"
/db_xref="GI:12323972"
/translation="MNAENQTTTHSKVISHVCTGTAKLGSVGPPIGLVDIGVSEV
AYIRVSLPGLEKQNDIKCEIQREGVCIQGVPIETAIIPSDTGCLYRMQVQQLCPG
PFSITFNLPGVDPLRFSNFRSDGIFEVVVKLGVRIPIS"
/20920..22411
/gene="F1466.5"
/join(<20920..20986,21073..21148,21227..21330,21407..>21472,
21739..21821,21911..21986,22064..22182,22264..22308,
22385..>22411)
/gene="F1466.5"
/join(20920..20986,21073..21148,21227..21330,21407..21472,
21739..21821,21911..21986,22064..22182,22264..22308,
22385..22411)
/gene="F1466.5"
/codon_start=1
/product="unknown protein; 20920-22411"
/protein_id="AAG51945.1"
/db_xref="GI:12323974"
/translation="MAAISPLWSSPQSFNSPRVITDTSRRSSISAISVLDSSNEQ
HRISRDHVKRRDVMQLQIASVFFLPIAISPAETINASEAFRVYTDENKFEIS
POVSTAIITGLGDFTRMESFGKVEAFATVSLGRDSWQKPVGVYAKLIDSRASKGY
YIEYTLQNGEARKHLYSAIGNATNGWYNRLYTVTGQFTDESAEQSSKIQKTVKFSR
FI"
/complement(23340..24808)
/gene="F1466.6"
/note="contains Pfam profile: PF00076 RNA recognition
motif (a.k.a. RRM, RBD, or RNP domain)"
/complement(join(<23340..23431,23548..23785,23879..>23919,
24008..24124,24234..24378,24602..>24808))
/gene="F1466.6"
/complement(join(23340..23431,23548..23785,23879..23919,
24008..24124,24234..24378,24602..24808))
/gene="F1466.6"
/codon_start=1
/product="putative RNA-binding protein; 24808-23340"
/protein_id="AAG51948.1"
/db_xref="GI:12323977"
/translation="MAYOPVPGSGPHLYNSPPGDTFTTKVFVGLAWETQSETLRQHF
EQYGEILEAVYADKNKTGSKGYGVETFDPEAABRACADPTPIIDGRANCNLSLG
RRPPLPYAVIPNMPDLPRLTLEMCRVLEHLLLEIRISNHFHTTSKELSIWMYAY
GPEYMSQSQSYVMGQYQLQVYGVPGVANSVPIYQSLQSLTPINGHQAQVQGS
VPGSHILQLGPTVSTMTTSSMPALQAPVPSGIPGAPVQSHIIVHSQFQWQSTASQ
TTR"
/complement(26257..27736)
/gene="F1466.7"
/note="similar to cinnamoyl-CoA reductase GB-CRA56103
[Bucalyptus gunnii]; contains non-consensus GG acceptor
splice site at exon 4"

```

```

complement(join(<26257..26401,26478..26667,26773..26935,
27062..27247,27332..27486,27593..>27736))
/gene="F1466.7"
/note="contains nonconsensus splice site"
complement(join(26257..26401,26478..26667,26773..26935,
27062..27247,27332..27486,27593..27707))
/gene="F1466.7"
/note="contains nonconsensus splice site"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 27707-26257"
/protein_id="AAG51951.1"
/db_xref="GI:12323980"
/translation="MAVAKQCVTGAGGFIASWLKVFLLSRGYVHGVTRVDPQEKND
HLRKLDNASKNLKLKADLFDDEGLFSAIDGSCGVFIASPVPEGEELERIPALTYIK
NLYEACTTKYQKVVVSVSIAAIVNPKPQDVAKDEDCDQTYLHLSLERYYLAKY
LTERALEWSKRNFADVTLCPSVIIGPRLQSTLNSSGLLKKFKIGKIKSLSDLEY
LVDVRDADALLLVENREATGRYICNSHSLVTDLSLMEKLMKMPKRNPFESFTEVKE
KEVRPLSAEKLKNLGKFRPLEETIDDSVVSVFEAAGDLPA"

```

Query Match	Score	DB	Length
Best Local Similarity	96.5%	Pred. No. 1.8e-83;	
Matches	557;	Conservative	0; Mismatches 19; Indels 1; Gaps 1;

```

QY 1 TGGGTTTTTATTCGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTTGGATAG 60
Db 39846 TGGGTTTTTATTCGATAACATGACAAATATTTATTTATTTTCATGAGTTTTTATTTGGATAG 39787
QY 61 CATCACAAATATTAATATATCATGCTGTTAATAACATGTTTTTGTTCCTTAAATACATCAATT 120
Db 39786 CATGACAAATATTAATATATATAGTGTGTTAATAACATGTTTTTGTTCCTTAAATACATCAATT 39727
QY 121 TTAATAATCAGACATTTGTTTTTAAAAATCAATCTAATCTCTATATATCACAACGACATTTGAC 180
Db 39726 TTAATAATCAGATATTTGTTTTTAAAAATCAATTTAAATCTCTATATCACAACGACATTTGAC 39667
QY 181 GGAATAATCAGGTAAAGAGAGAAAATAAGAAATGAGAGATAGAGAGATTTCTATGGAATAA 240
Db 39666 GGAGAAATTCAGGTAAAGAGAGAAAATAAGAAATGAGAGATAGAGAGATTTCTATGGAATAA 39607
QY 241 AGAAAGAGAGAACATGTAGTGAAC-AAAAATAAGAGATATGATATATATTTTATGAGA 299
Db 39606 AGAAAGAGAGAACATGTGGGTGAACAAAAATAAGAGATATGATATATATTTTATGAGA 39547
QY 300 GGTGGTGAAGATTTATTTAGAGAGGGAGAGAGAAATAGAAAAGAAAAAAGACATGGTGA 359
Db 39546 GGTGGTGAAGATTTATTTAGAGAGGGAGAGAGAAATAGAAAAGAAAAAAGACATGGTGA 39487
QY 360 ATCTGAAGAAGATCAATGTTGTTAAAGATGAAGAGAGAAAAGAGAACTCCATGGCTAAAGT 419
Db 39486 ATCTGAAGAAGATGAATGTTGTTAAAGATGAAGAGAGAAAAGAGAACTATATGCTAAAGT 39427
QY 420 CTCGTAAGAAGATGAAAAGAAACAAAAGAGGAAGAGAAAGAGAAAGCGCTAAATATAG 479
Db 39426 CTCGTAAGAAGATGAAAAGAAATAAAGAGGAAGAGAAAGAGAAAGCGCTAAATATAG 39367
QY 480 ACTAACTATTGCCAAAATTTCTGTAGCCGACATACTATTGTCCTCAAGTTATTTTGT 539
Db 39366 ACTAAATATTGCCAAAATTTCTGTAAACCACAGTACTATTGTTGTTCTAAAGTTATTTTGT 39307
QY 540 GTATCTTTTCAAGTCAAAAGTTATTTCTTACATATA 576
Db 39306 GTATCTTTTCAAGTCAAAAGTTATTTCTTACATATA 39270

```

```

RESULT 6
ATT32A17/c 104386 bp DNA linear PLN 23-MAR-2001
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone T32A17 (ESSA
project).
ACCESSION AL161813 AC007635
VERSION AL161813.1 GI:7321058
KEYWORDS thale cress.
SOURCE

```


[illegible]

<p>Arabidopsis thaliana, PIR2:T05898</p> <p>Contains Prokaryotic membrane lipoprotein lipid attachment site AA12-22:Protein kinases signatures and profile AA763-784:Protein kinases signatures and profile AA883-895</p> <p>contains EST gb:F14412, Z17737, R84191, T22574, AA394763, AA395690, F14413*</p>									
Query Match	29.4%	Score 499.8;	DB 8;	Length 179771;					
Best Local Similarity	90.6%;	Pred. No. 1.7e-77;							
Matches 557;	Conservative 0;	Mismatches 52;	Indels 6;	Gaps 2;					
QY 1	TCGGTTTTATTCGGATACATGCACAATATTTATTTATTTTCATGAGTTTTTATTGGATAG	60							
Db 17646	TCGGTTTTATTCGGATACATGCATATATTTATTTATTTTCATGAGTTTTTTCGGATAG	17587							
QY 61	CATGCAAAATATTAATATATCATGCTGTTAAATAACATGTTTGTCTTTAAAATACATGCATTT	120							
Db 17586	CATGCAAAATATTAATATATCATGCTGTTAAATACATGTTTGTCTTTAAAATACATGCATTT	17527							
QY 121	TTAAAATCAGACATTTGTTTTTAAAATCAATCTTAATCTCTATATATCACACGACATTCGAC	180							
Db 17526	TTAAAATCAGACATTTGTTTTTAAATCAATCTTAATCTCTATATATCACACGACATTCGAT	17467							
QY 181	CGAAATTCAGCTAAAAGAGAAATAAAGAATGAGAGATAGAGAGATTTCTATGCAAAA	240							
Db 17466	GGAGAAATTCAGGCAAAAAGAGAAACAAGATGAGAGATAGAGAGATTTCTATGCAAAA	17407							
QY 241	AGAAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGATATATTTATGAGAG	300							
Db 17406	ATAAAGAGGGAACATGTGGGTGAACAAAATAAAGAGATATGATGATATATTTATGAGAG	17347							
QY 301	GTGGTCAACATTTATTTT-----TAGSAGGGGAGAGAGAAATAGAAAAGAAAATGACATG	355							
Db 17346	GTGGTCAACATTTATTTT-----TAGSAGGGGAGAGAGAAATAGAAAAGAAAATGACATG	17287							
QY 356	GTGAATCTCGAAGAAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAGAACTCCATGGCTA	415							
Db 17286	GTGAATCTCGAAGAAGATGAATTTGTTTAAAGATGAAGAGAGAAAGAGAACTCCATAGCTA	17227							
QY 416	AAGTCTCTGAAGAGATGAAAAAGAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGCTAAA	475							
Db 17226	AAGTCTCTGAAGAGATGAAAAAGAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGCTAAA	17167							
QY 476	ATAGACTAACTATTGCCAAAATTTCTGTAGCGGCAAAATACTATTTGGTCCAAAGGTTATT	535							
Db 17166	ATAGACTAA-TATTGCCAAAATTTCTGTAGTACAGAGTACTATTTGTTCCAAAGGTTATT	17108							
QY 536	TTGCTGATTTCTTTTGAAGTCAAAAGTATTTCTTTACATATATCTCTAAAATATAGCCGAT	595							
Db 17107	TTGCTGATTTCTTTTCAAGTCAAAAGGTTATTTTGTATATATATCTCTTTTATAAAAAAT	17048							
QY 596	ACCAATTTTTTCCACA 610								
Db 17047	ATAACGTTTTCGAAA 17033								
RESULT 8									
AF071527									
LOCUS	AF071527	95519 bp	DNA	linear	PLN 06-MAY-1999				
DEFINITION	Arabidopsis thaliana BAC F9H3, from chromosome IV near 18.8 cm, complete sequence.								
ACCESSION	AF071527								
VERSION	AF071527.1	GI:3236479							
KEYWORDS	HTG.								
SOURCE	Arabidopsis thaliana.								
ORGANISM	Arabidopsis thaliana								
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 95519)								
REFERENCE	Huang, E.N., Parnell, J.I.D., de la Bastide, M., Schutz, K., Habermann, K., Dedhia, N.N. and McCombie, W.R.								
AUTHORS									

Genomic sequence of Arabidopsis thaliana BAC F9H3, chromosome IV, 18.8 cM

Unpublished

2 (bases 1 to 95519)

Huang, E.N., Parnell, L.D., de la Bastide, M., Schutz, K., Habermann, K., Dedhia, N.N. and McCombie, W.R.

Direct Submission

Submitted (10-JUN-1998) Cold Spring Harbor Laboratories, Lita Annenberg Hazen Genome Center, 1, Bungtown Road, Cold Spring Harbor, NY 11724, USA

3 (bases 1 to 95519)

Parnell, L.D. and McCombie, W.R.

Direct Submission

Submitted (31-DEC-1998) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

Arabidopsis thaliana BAC F9H3 from chromosome 4, near 18.8 cM

4 (bases 1 to 95519)

Parnell, L.D. and McCombie, W.R.

Direct Submission

Submitted (01-FEB-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.5 cM

5 (bases 1 to 95519)

Parnell, L.D.

Direct Submission

Submitted (06-MAY-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724

Arabidopsis thaliana BAC F9H3 from chromosome IV near 18.8 cM

BAC F9H3 is assigned between YACs CIC4A7 and CIC8B1 and maps to near 18.8 cM on the Lister & Dean RI map. Position 1 of F9H3 is oriented toward the centromere and position 95519 is oriented toward the telomere. For more information on the mapping, sequencing and annotation of F9H3, please see <http://www.cshl.org/arabweb/F9H3-titlepage.html>. A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by GenScan (<http://ccr-081.mit.edu/GENSCAN.html>), MZPF (<http://www.cshl.org/genefinder>) and GRAIL (<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites predicted by NetPlantGene (<http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>). Alternate exons not used in building the gene models are presented on the web pages associated with F9H3. Genes are numbered according to the scheme BAC gene number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (<http://expasy.hughe.ch/sprot/prosite.html>) and Pfam (<http://pfam.wustl.edu/>) libraries. A description of these categories can be found at <http://muntjac.mips.biochem.mpg.de/arabi/>. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F9H3, please direct email to Larry Parnell at parnell@cshl.org.

BAC F9H3 was sequenced as part of the Arabidopsis genome sequencing effort of the Cold Spring Harbor Consortium. For additional information, please see <http://www.cshl.org/arabweb>. Fingerprint data indicate F9H3 overlaps with F4C21 toward the telomere and with T5L23 toward the centromere. F9H3 contains marker ml233 at 18.8 cM on the Lister & Dean RI map.

1. 95519
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F9H3, please direct email to Larry Farnell at parnell@cslh.org. BAC F9H3 was sequenced as part of the arabidopsis genome sequencing effort of the Cold Spring Harbor Consortium. For additional information, please see <http://www.cslh.org/arabweb>. Fingerprint data indicate F9H3 overlaps with F4C21 toward the telomere and with T5L23 toward the centromere. F9H3 contains markers ml233 at 18.8 cM on the Lister & Dean R1 map.

```

Location/Qualifiers
1. 95519
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"

```

```

/chromosome="iv"
/map="18.8 cM"
/clone="BAC F9H3"
/note="hybridizing YAC is C1C8B1"
misc_feature
1..31783
/note="overlap with T5L23, GenBank accession number
AC005142, from position 26 to 31808"
851..1719
/gene="F9H3.20"
/pseudo
/evidence-not_experimental
complement(981..1120)
/note="function-unclassified; similar to F6N23, GenBank
accession number AF058919"
rpt_type-dispersed
3173..3969
/note="function-putative_transposon; similar to FCAL,
GenBank accession number Z97336; moderately similar to
reverse transcriptase gene of DW15 non-LTR
retrotransposon, GenBank accession number L47193"
rpt_type-dispersed
5889..6303
/note="function-chloroplast_repeat"
rpt_type-dispersed
complement(6653..12091)
/gene="F9H3.19"
/note="encodes putative calcium channel; genomic copy of
EST W43766; genomic copy of EST T43251; EST T43251
apparently encodes intronic sequence; possible alternate
exon 1 from position 12772 to 12846 is predicted by a
NetPlantGene splice donor but has no EST coverage; gene
model last edited on 27 Jan 99"
/evidence-not_experimental
complement(join(6653..6746,6833..6870,6991..7062,
7175..7261,7436..7526,7682..7746,7985..8153,8229..8292,
8380..8487,8757..8814,8891..9034,9127..9192,9311..9383,
9530..9591,9745..9822,9929..10034,10127..10293,
10391..10451,10557..10580,10644..10774,10890..10930,
11292..11403,11500..11643,11972..12091))
/gene="F9H3.19"
/note="functional catalog ID=07.01"
/codon_start=1
/evidence-not_experimental
/product="putative calcium channel"
/protein_id="AAD11598.1"
/db_xref="GI:4206210"
/translation="MEDFLGRDLSGGGTRVRSEAIHGTTPQKAAALVDLAEDG
IGLVEILDQSFGESARYIFTLRLDILNSLYPALFLNFFEQPLCEKNPKPSCK
DRDYVYLGELPYLTNAESIIEASSRIEFTSLNLKVKACVILFVDVLDVLYLSPL
AFDFLPFIAPVRVYIFILSIQRQCADHVMELRDLVLLSGMLCTYUNILALWMLF
LFASIAVMFEEDTQGLTVFTSYGATLYQMFLFTTSNPDVMTIPAKKSRWSVFE
VLYVLGVFTVNLILAVYDSFEKQLAKQVSGDOMKRMLEKAFGLIDSQKGEID
KQCKILFEQLNRYRTLPIKSKKEEFLGFLIDELDDTRPKINKDEFADICQAIARFQK
EVPSLFEPFOIYHSLSQOLRAEVRSPNFYALISFLIINFIAVVVETTLDEES
AOKPQVAFVFGVITVLEMAKLIYTYGFWREGANRFDLVTWVIVIGETATFIT
PDENFTFNGEWIRYLLRLMLRILRLMNVQRYRAFATITLIPSLPVLGTIFCV
LCYCSIGVQVFGGLVGNAGNKKLPELAEDDYLLENFDYPNGMVTLLNLLVMNQ
VMWESYKDLTGTWMSITYVFSVYITILLNLVAVFLEAFFTELDEEBEKKQGGQD
SQEKNRRRSAGSKRSORVDTLHHMLMGDELSKPECSTSDT"
complement(13149..18831)
/gene="F9H3.18"
/note="encodes putative glucan synthase component; gene
model last edited on 5 Jan 99"
/evidence-not_experimental
complement(join(13149..14597,14778..16800,16961..18831))
/gene="F9H3.18"
/note="similar to 1,3-beta glucan synthase; functional
catalog ID=01.05.99; functional catalog ID=09.01"
/codon_start=1
/evidence-not_experimental
/product="putative glucan synthase component"
/protein_id="AAD11597.1"
/db_xref="GI:4206209"

/translation="MSLRHRTVPQGRPLAAEAVGIEEPEYNIIPVNNLLADHPSLR
FPEVRAAAALKTVGDLRRPPYVOMRSHYDLDLWLLALFFGQKDNVNRQREHVMVHLHA
NAQRLSPDPNDISLDSAVRRFRKLLANYSSW:SYLCKNINISDRNPSRREL
LYGLYLLIMGEANRLMEPECICYIFHNWASELNKILIEDCLNDENQGPYLPISGEN
AFGTGVKPYDIYTAETDSKNGTVAHCKRNVDINEYFWTDRCFSKLKPDLGLGS
NFFKSRGKSVGKTGFVERRTFFLYRSFDRLWMLALFOAAIIVAWEEKPDTSSVTR
OLNNALKARDVOVRLTLVLTWSGMRLLLOALDAOSYPLVSRBETHFRFRLMKLVIA
AMNWIATVLYINWIKQKQDQKQWNSAAITFIYOFIYAGAFLEPILALALFIIPW
MNFLEETNMKIFIALTWFGQKSFVGRGLRGLVDNIKYSFTWIFVIALATFTFSYFL
QVPMKPSKLLMNLKDVQWYEHQFQGSNRF:SVAILMLPEVLVIYIMDIQIYAIYSS
IVGAVVGLFDHLGIRDGQRLRFQFASAIQFNLMPEQLLNARFGNFKDGLHHR
LKLYRGRFPFKLESNOEANKFALINWEI:WAFREEDIVSDREVEELPKNSWDV
TIWPCFLCINELLALSOARELIDAPDKW:WFKICKNEYRCACVVEADSLKHLL
SLIKVDTEHSILTVFQIINOSIQSOFTK:PRVDLLPKIYETLQKLGLVNDDETD
SGRVNVLQSLYEIATFQIEKTEITQLSNLGLTPRDPASKLQFNALRUPDASND
FYRVNRRLHILTSRDSMHSVPVNLAEARRIAFNSLPHNMHPAOPKUMAFSVLT
PYSEEVYKQELRNEDTIGTLYLYQTYADEVKNFKRMHREGIKTDSBLWTK
LRDLRLWASYRGOTLARTVRGMYRYRALKMLAFLDASEMDIRGAQELGYSQKALQ
ELGQSDGFVSENDRSLSRASSSVLYKGRVYALMKFTYVACQIYGSQAKKE
POAEIILYLMKNEALRIAYDVEPAGRGETUYSLVKYDHOLEKEVEIFRVKLPQP
VKLGECPNQNHAMIFTRDAVQTDIMNQDSIFELAKMRNLLOEYNNHYHGRKPTI
LGVREHIFTGVSLSLAWFMSAQETSFVTLGQFVLANPLKVRMHYGHDPDPRFWFLSR
GGISKASRVINISDIFAGNCTLRGNVTHFEYIQVGRDVLGMLQISMFEAKVASG
NGEQLSRDVRYLGRDLDFRMLSFYFTVGI:FN:MMVILTVYAFIWRVYIALSGV
EKSALADSTDTNALGVLNQOFLIQLGLFTALPM:VWLSLEEGFLLAIWNTFRMOIQ
LSAVYTFSGMTRAHVFGRTLHGGAKYRAT:GCVVVEHKGFTEYVRLYARSHFYRAI
ELGLILIVYASHSPIAKDSLIYIATMTISNF:VISIMAPFVNPSPGDFWLKTVDFE
DFMMWITQGRISTKSEKWEKWEYEDHLLNIGKAGLEVEILLVLURFFFOGIVY
OLKTANGSTSLFVFLFYFAIFVFLFVIC:YARDKYSKAHRYRLVQFLVILAI
LVIIVALLEFTFSFDITSLIAFIPGWIILIAQTORKNKLYITFWNAVYSVARM
YDILFGLIMVPAFLSMWPGFQSMQTRILENCAFSRGLRIMOIVTQKSKGDV"
complement(20861..22302)
/gene="F9H3.17"
/note="encodes hypothetical protein; gene model last
edited on 5 Jan 99"
/evidence-not_experimental
complement(join(20861..21094,21437..22039,22145..22302))
/gene="F9H3.17"
/note="similar to F21B7.5, GenBank accession number
AC002560; functional catalog ID=19"

Query Match 29.2%; Score 497; DB 8; length 95519;
Best Local Similarity 92.5%; Pred. No. 5.8e-77;
Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

Qy 1 TGGGTTTTTATGGATACATGACAAATATATTATTATTTTCATAGTTTATTATTGGATAG 60
Db 23303 TGAGTTTTTATGGATACATGACAAATATATTATATATTATATGATTTTGTGGATAG 23362

Qy 61 CATGACAAATATTAATATATCAGTGTAAATAACATGTTTCTTCTTAAATACATGCATT 120
Db 23363 CATGACAAATATTAATATATCAGTGTAAATAACATGTTTCTTCTTAAATACATGCATT 23422

Qy 121 TTAATAATCAGACATTGTTGTTTAAATAACAAATCTAATCTCTTATTCACACGACATTGAC 180
Db 23423 TTAATAATCAGACATTGTTGTTTAAATAACAAATCTAATCTCTTATTCACACGACATTGAT 23482

Qy 181 GGAATAATTCAGGTAAAAGAGAAATAAAGAAATGAGATAGAGAGATTTCCTATGGAATA 240
Db 23483 GAAGAAATTCAGGCAAAAAGAGAAACAAAGAAATGAAAGATAGAGAGATTTCCTATGGAATA 23542

Qy 241 AGAAGAGAGACATGTAGGTGAACAAATAAAGAGATATGATATATATTTATGAGAG 300
Db 23543 AGAAGAGAGAACATGTGGGTGAACAAATAAAGAGATATGATATATATTTATGAGAG 23602

Qy 301 GTGGTGAAGATTATTTAGGAGAGGGAGAGAAATAGAAATAGAAAAGAAATACACATGGTGA 360
Db 23603 GTGGTGAAGA-TATTTAGGAGAGGGAGAGAAAGTAGAAGAAAGAAATACACATGGTGA 23661

Qy 361 TCTGAAGAAGATGAATTTGTGTTAAAGATGAGAGAGAAAGAGAACTCCATCGCTAAAGTC 420
Db 23662 TTTGAAGAAGATGAATTTGTGTTAAAGATGAGAGAGAAAGAGAACTTCATGGCTAAAGTT 23721

```


lemcke@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bbsrc.ac.uk
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with ATCHRIV8 at the 5' end and an
 overlap with ATCHRIV10 at the 3' end.

FEATURES

source	Location/Qualifiers
gene	1..159629 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 6741..9588 /gene="AT4g03480"
gene	complement(join(6741..7064,7161..7708,7783..8101, 8409..8866,9283..9343,9445..9588)) /gene="AT4g03480"
CDS	complement(join(6741..7064,7161..7708,7783..8101, 8409..8866,9283..9343,9445..9588)) /gene="AT4g03480"
	/note="contains similarity to mammalian ankyrins similarity to similar to A. thaliana hypothetical protein F21B7.8, GenBank accession number AC002560 Contains ATP/GTP-binding site motif A (P-loop) AA297-304 contains EST gb:H37081, T42092" /codon_start=1 /product="hypothetical protein" /protein_id="CAB77833.1" /db_xref="GI:7270671"
exon	complement(6741..7064) /gene="AT4g03480"
intron	complement(7065..7160) /gene="AT4g03480"
exon	complement(7161..7708) /gene="AT4g03480"
intron	complement(7709..7782) /gene="AT4g03480"
exon	complement(7783..8101) /gene="AT4g03480"
intron	complement(8102..8408) /gene="AT4g03480"
exon	complement(8409..8866) /gene="AT4g03480"
intron	complement(8867..9282) /gene="AT4g03480"
exon	complement(9283..9343) /gene="AT4g03480"
intron	complement(9344..9444) /gene="AT4g03480"
	/number=5
exon	complement(9445..9588) /gene="AT4g03480"
gene	10335..13501 /gene="AT4g03490"
gene	complement(join(10335..10525,10983..11078,11590..12133, 12345..12500,12627..12772,12871..13501)) /gene="AT4g03490"
CDS	complement(join(10335..10525,10983..11078,11590..12133, 12345..12500,12627..12772,12871..13501)) /gene="AT4g03490"
	/note="contains similarity to mammalian ankyrins similarity to similar to A. thaliana hypothetical protein F21B7.8, GenBank accession number AC002560" /codon_start=1 /product="hypothetical protein" /protein_id="CAB77834.1" /db_xref="GI:7270672"
exon	complement(10335..10525) /gene="AT4g03490"
intron	complement(10526..10982) /gene="AT4g03490"
exon	complement(10983..11078) /gene="AT4g03490"
intron	complement(11079..11589) /gene="AT4g03490"
exon	complement(11590..12133) /gene="AT4g03490"
intron	complement(12134..12344) /gene="AT4g03490"
exon	complement(12345..12500) /gene="AT4g03490"
intron	complement(12501..12626) /gene="AT4g03490"
exon	complement(12627..12772) /gene="AT4g03490"
intron	complement(12773..12870) /gene="AT4g03490"
exon	complement(12871..13501) /gene="AT4g03490"
misc_feature	13514..15185 /note="marker mi233"
gene	14170..17288 /gene="AT4g03500"
CDS	join(14170..14204,15103..15848,15939..16477,16650..17288) /gene="AT4g03500"
	/note="hypothetical protein with some similarity to ankyrin similarity to" /codon_start=1 /product="hypothetical protein"

```
/protein_id="CAB77835.1"
/db_xref="GI:7270673"
/translation="MKNESRTSEIKLIFSGVFSKGRPDVHLDRRVAENHQAIPIR
NNVSTPSLDLSTLFDSETSEKPNMDPTMAARAGKENYLRNSNYSLSVAPTILVNDRG
DIICHLAASGHVSLVRYIIQKCPGLLLKSNMGEVALHAAEAGHLDVVMNIDFIN
NTSTNLPAVKRIYFAKNKQDIALHVALKGHEVAALYSAAKSLSFVNRDGFSP
LYLAIEAGHTSLVTVMGHTNELSSKVGGRSIVHAAKANKRDKIDALLSKDASLNL
RDEGRTSLSFASGYQGFSLYLPDKNRDVIYVSDDLGFLPTHMAKYGHVQILEIL
KHCPEATELDRDQNLHLAAKYGKLVKIFILSCCKDKKKLLNEQDVNGNTPLH
LATINHPKVVMETWDRHVDLKRNYIGFTALDVAEENIDSSIVHORUTWALLINA
GAPKSSPTENLRSKPDGKYKDRVNTMLVATLVATVMTAGTFLPGVYNDSEFP
HLGMAVAKRTAFQVFLVCDTLAMYSIIITVALIWAQGLDLSILKAFNALPFLGL
ALTSMSIAFMAGTVAYSHLPILGIVGLGICIFLLVLLLLVYVSPYAHQAQPLLHR
IFYYPYFLKLLAAGDNKNIVDVYTSDE"
14170..14204
/gene="AT4g03500"
/number=1
14205..15102
/gene="AT4g03500"
/number=1
15103..15848
/gene="AT4g03500"
/number=2
15849..15938
/gene="AT4g03500"

exon
intron

Query Match 29.28; Score 497; DB 8; Length 159629;
Best Local Similarity 92.58; Pred. No. 5.3e-77;
Matches 545; Conservative 0; Mismatches 40; Indels 4; Gaps 2;

QY 1 TGGGTTTTTATGGATAACATGACAAATATTTATTTTTCATGAGTTTTTATGGATAG 60
Db 29758 TGAGTTTTTATGGATAACATGACAAATATTTATATATCATCAATTTTGTGGATAG 29699

QY 61 CATGACAAATATATATATCATGTTTAATACATGTTTTGTTCTTAAATACATGCATT 120
Db 29698 CATGCAAAATATATATATCATGTTTAATACATGTTTTGTTCTTAAATACATGCATT 29639

QY 121 TTAATCAGACATTTGTTTAAATCAAACTAATCTTATATATCAACAGCATTGAC 180
Db 29638 TTAATCAGACATTTGTTTAAATCAAACTAATCTTATATCAACAGCATTGAT 29579

QY 181 GGAATTCAGGTAAAAGAGAAAATAAGAAATGAGAGATAGAGATTTCTATGGAAA 240
Db 29578 GAAGAATTCAGGCAAAAAGAGAAAACAAGAAATCAAGATAGAGATTTCTATGGAAA 29519

QY 241 AGAAGAGAGAACATGTAGGTGTAACAAATAAGAGATATCATGATATATTTATGAGAG 300
Db 29518 AGAAGAGAGAACATGTGGGTGTAACAAATAAGAGATATCATGATATATTTATGAGAG 29459

QY 301 GTGGTGAAGATTATTTAGGAGGGGAGAGAAATAGAAAAGAAAATGACATGGTGAA 360
Db 29458 GTGGTGAAGA-TATCTAGGAGGGGAGAGAAAAGTAGAAAAGAAAATGACATGGTGAA 29400

QY 361 TCTGAAGAGATGAATTGTGTTAAGATGAGAGAGAAAGAGAACTCCATGGCTAAAGTC 420
Db 29399 TTTGAAGAAGATGAATTGTGTTAAGATGAGAGAGAAAGAGAACTTCATGGCTAAAGTT 29340

QY 421 TCGTAAAGAAGATGAAAAGAGAAACAAAGGAAGGAAGAGAAAGAGAAAGCTTAAATAGA 480
Db 29339 TCGTAAAGAAGATGAAAAGAGAAAGAA---AAGAGAGAAAGAGAAAGAAATAAATAGA 29283

QY 481 CTAACTATPGCCAAATTTCTGAGCCGACAAATACCTATTTTGGTCCAAAGTTATTTGTG 540
Db 29282 CTAAATATGCAAAATTTCTGAGCAGCAAGTACTATTTTGGTCCAAAGTTATTTGTG 29223

QY 541 TATCTTTTGAAGTCAAAAGTATTTCTTACATATACTCTTAAATAATA 589
Db 29222 TATGTTTTCAAGTCAAAAGTATTTTGTATATATACTCTTATATATA 29174

RESULT 11
ATH250185
```

LOCUS ATH250185 369 bp mRNA linear PLN 15-AUG-2001
DEFINITION Arabidopsis thaliana mRNA for NIMIN-2 protein (nimin-2 gene).
ACCESSION AJ250185
VERSION AJ250185.1 GI:12057155
KEYWORDS nimin-2 gene; NIMIN-2 protein.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 369)
Weigel, R., Bauscher, C., Pfitzner, A.J., and Pfitzner, U.M.
NIMIN-1, NIMIN-2 and NIMIN-3, members of a novel family of proteins from Arabidopsis that interact with NPR1/NIM1, a key regulator of systemic acquired resistance in plants
Plant Mol. Biol. 46 (2), 143-160 (2001)
JOURNAL MEDLINE 2135044
PUBMED 11442055
REFERENCE 2 (bases 1 to 369)
Pfitzner, U.M.
Direct Submission
Submitted (07-OCT-1999) Pfitzner U.M., General Virology, Institute of Genetics, Emil Wolff Str. 14, 70599 Stuttgart, GERMANY
JOURNAL Related sequence: AB023041.
COMMENT Location/Qualifiers
FEATURES
source
1..369
/organism="Arabidopsis thaliana"
/cultivar="Col-0"
/db_xref="taxon:3702"
/tissue_type="green vegetative tissue"
/dev_stage="3-wk-old"
1..369
/gene="nimin-2"
1..369
/gene="nimin-2"
/function="putative role in SAR"
/codon_start=1
/evidence="experimental"
/product="NIMIN-2 protein"
/protein_id="CAC19845.1"
/db_xref="GI:12057156"
/db_xref="SPTREMBL:Q9LUJ3"
/translation="MNNSLKKEERVEEDNGKSDGNRGKPSPEVTVTTEBVDPEFKI
LRRVHVATRTVAKVNGVAGEGLPSKSKRSCNGLRNSLDCNVGRGEFDEINRVGL
QGLGLDLACKPEPDSVLSL"
BASE COUNT 113 a 53 c 121 g 82 t
ORIGIN
Query Match 21.7%; Score 369; DB 8; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.5e-54;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 863 ATGAACAACCTCTTTGAAGAAAGAACCGGTAGAGAGATAACGGAAATCTCAGCGT 922
Db 1 ATGAACAACCTCTTTGAAGAAAGAACCGGTAGAGAGATAACGGAAATCTCAGCGT 60
QY 923 AACAGAGGAAACCGTCGACGAACTGTTCCGACGGTAACGAGAGAGGTGATGAG 982
Db 61 ACAGAGGAAACCGTCGACGAACTGTTCCGACGGTAACGAGAGAGGTGATGAG 120
QY 983 TTTTCAAGATATTACGAGAGATACACGTGGCGACACGAAACCGTTCGAAAGCTTAACGCG 1042
Db 121 TTTTCAAGATATTACGAGAGATACACGTGGCGACACGAAACCGTTCGAAAGCTTAACGCG 180
QY 1043 GGTGTTGCTGAGGAGAGTTACCGTCTAAGAGAGAGAAACGAGATCAGAAATCTCGGTTG 1102
Db 181 GGTGTTGCTGAGGAGAGTTACCGTCTAAGAGAGAGAAACGAGATCAGAAATCTCGGTTG 240
QY 1103 AGAAACTCGTTGGATTGTAACGGCTTCGAGACGAGAAATTCGATAGATTAATCGGTC 1162
Db 241 AGAAACTCGTTGGATTGTAACGGCTTCGAGACGAGAAATTCGATAGATTAATCGGTC 300
QY 1163 GGGTTACAGGGTTTGGGTTTGGATCTGAACCTAAACCGGACCAAGACAGCGTTAGTTTA 1222

Db 301 GGGTTACAGGTTGGTTGAGTCACTGAACGTAAACCGGAACAGACAGCGTTAGTTTA 360
 QY 1223 TCGTTGTAG 1231
 Db 361 TCGTTGTAG 369
 RESULT 12
 AC007203/c
 LOCUS
 DEFINITION Arabidopsis thaliana chromosome I BAC T10P12 genomic sequence,
 complete sequence.
 ACCESSION AC007203
 VERSION AC007203.3 GI:4996503
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.S.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Unpublished
 2 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (03-APR-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 3 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (04-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 4 (bases 1 to 95190)
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
 Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P.,
 Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
 Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
 Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (16-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 On Jun 4, 1999 this sequence version replaced gi:4678187.
 Bases 1-6718 of clone T10P12 overlap with bases 122,554-129,271 of
 'IGF' BAC clone F1121, gb|AC005687
 e-mail for correspondence: arab@sequence.stanford.edu
 Genes with similarity to proteins in the databases are described as
 'putative', '-like' or 'similar to'. Genes that have EST
 similarity but no significant protein similarity are described as
 'unknown proteins'. Genes that are annotated based only on gene
 prediction software are described as 'hypothetical proteins'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://combio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
 http://genome.stanford.edu/~chris/GENSCAN.html), Fexa (V.Solovyev
 & A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/),
 NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of

Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>) and
 eMotif (Newell-Manning, C.G. Wu, T.D. & Brutlag, D.L.,
 http://motif.stanford.edu/projects.html).

FEATURES

source

Location/Qualifiers
 1..95190
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="I"
 /clone="T10P12"

misc_feature

1..6718
 /note="overlap with bases 122,554-129,271 of 'IGF' BAC
 clone F1121, gb|AC005687"
 9474..10628

repeat_region

complement(19003..23764)
 /gene="T10P12.1"

gene

complement(19003..23764)
 21074..21678,21933..22099,22427..22640,22931..23131,
 23362..23764)
 /gene="T10P12.1"

/note="Similar to reverse transcriptase: Similar to reverse
 transcriptase, part of putative retrotransposon"

/codon_start=1

/protein_id="AAD39270.1"
 /db_xref="GI:5080760"

/translation="MDMIFINVGVKVRKIKNSKTAACAWGTLDLILYLVKSLPNRYIL
 QLVYNYRMQNSKTLEENVEQFMISDLNNIQVDPDEVAQAAILLSALPDSCDM
 LKETLYGREGIKLDIVISAALSKELELDSEVNGESKGDHGEWIIDTECSFHM
 TPKEVHIDFEAKSGKVRMPNPSFLEVGKIGKVRFTNQDGTSCFGNDISIKIKCFNF
 VVGKTPRTSFGTAQHVTKKLDYVHSDLWSSNPVYSI SKQYFTSFDDNSRKWVIN
 FUNKDEATPEELTWSVIPSLGRKRFGCIVYVSHQECKLDPKAKKGVFVGPSSHVD
 PSSAENQTRQIARDPRRQIVLPRLKNDYDEEVLDEIAGYAVLTEDGNGPPAS
 TEEENQTRQIARDPRRQIVLPRLKNDYDEEVLDEIAGYAVLTEDGNGPPAS
 FOEALQSDSGKWLAADEIQSLINKTIVLERNKSKPGICCKVFKRKGAGIAGE
 KPRKARLVTKGYSOKEGIYFOEISPVVYVSIIRLLSIVHDLMEQLQMDKFTAFLL
 HGYLDEIMPDQEGVYHEKYPKRYIYLLVYVDDILIAASKDKKHVCEKLVLLNSFEM
 KOLGRKAKILGMEIIRKREGGLTISOEYLILKLPATVEYVQKQKVMRTVTPYRSVA
 GSLMYSMIGTRPDLAHSVGLCMWRKPLKDHQAVKWLRCIVGTMRKLCYKNEGG
 LALGYCDSYAAADKEKRSTVFTGGNTISWRSSLOKVVALUSKRAEYMAUTDA
 TNETVWLKGLVNELGFTQSVNIHCDQSQSIADAKNAVYHRTKHIDYKHFIRLELA
 ANNELTKLEGGSLEEEKKQVLAWEKEDDAATYNNKLVVELKELKIPVFPNLLIP
 NSLGNRSAGELADGVTDHASSNLRDEDMCPSEEPAP"
 29564..30696

complement(31637..35319)
 /gene="T10P12.2"

repeat_region

gene

complement(31637..32634,33345..33502,33679..33916,
 34104..34307,34383..34539,34828..35319)
 /gene="T10P12.2"

CDS

/note="Similar to reverse transcriptases; protein is
 flanked by LTRs, potential retrotransposon"

/codon_start=1

/protein_id="AAD39271.1"
 /db_xref="GI:5080761"

/translation="MMEQVQRIQTGYFSGGTSPDEASWRSVRNFSSKCSVEYLV
 DIAVHFLEGDHLWRSVTKARROADFWAFENAKYCPQEALEDMARFELTQ
 GERSVREYDREFNLLVYAGRMEDDQAOMLRGLRDPDRVRCRVSYATKAALVE
 TAAELMWAVVQPAVQAGVQGVQOQVQOQLAHGACAPQGYTTRIGSTSNRAITGL
 LAHEVCVGTLLVDGVHVLDFSPKASRGNIRGDPGEQSGSVYKVGQGF
 LEVIGRAKVVYAEKMEKCEAYLMTISMPESGVQVANSIRVDEFEVDVFSLOGLP
 PWSNDPFTIVLEPFTALLSKHPTQLQWIDELLQDLRGATCFSKIDLTSGYHOIPI
 AEADIRKTAFTKRGVFWIIPFARAEALDAWIVDRLTKSAHLAIIRKTDGAVALT
 KKVSVETKLHGVPVSIIVSDRSKFTSAFRAFOAEMGTVMQSTAYHPQTDQSERT
 IQILEDMLRMCVLDWGHADQLSLIEPAYNNYKQASIGMTPEVLYGRPCQTOLCMT
 QGERSYIGANYVQETTERIGVLLKMKKEVQQRHRSYADKRRRELEFEFEVGVYKLEMA
 MLRGNRSISETKLSPRYGPFRVYVAYRDELTDVNRFAHFHVFSVMRLKCLH
 KDELVAKIPEDLQPNNTLEARPGRVLRERIKELRKKIPLIKVLDCDCDGVTEETWEP
 EVMKARF"
 35970..37136

repeat_region

gene

37249..42330
 /gene="T10P12.3"

CDS

join(37249..37347,37714..38098,38265..38532,38897..39097,
 39215..39508,39603..40889,41069..42330)
 /gene="T10P12.3"

/protein_id="AAD39274.1"

[illegible]

* 34188 50517: contig of 14330 bp in length
* 50518 50617: gap of 100 bp
* 50618 65071: contig of 14454 bp in length
* 65072 65171: gap of 100 bp
* 65172 81054: contig of 15883 bp in length
* 81055 81154: gap of 100 bp
* 81155 95084: contig of 13930 bp in length
* 95085 95184: gap of 100 bp
* 95185 133197: contig of 38013 bp in length
* 133198 133297: gap of 100 bp
* 133298 171763: contig of 38466 bp in length.

FEATURES

source
1..171763
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RPC1-24 Male Mouse BAC"
1..1159
/note="assembly_fragment"
1260..2312
/note="assembly_fragment"
2413..3453
/note="assembly_fragment"
3554..5119
/note="assembly_fragment"
5220..7819
/note="assembly_fragment"
7920..9949
/note="assembly_fragment"
10050..14159
/note="assembly_fragment"
14260..18550
/note="assembly_fragment"
18651..25307
/note="assembly_fragment"
25408..34087
/note="assembly_fragment"
34188..50517
/note="assembly_fragment"
50618..65071
/note="assembly_fragment"
65172..81054
/note="assembly_fragment"
81155..95084
/note="assembly_fragment"
95185..133197
/note="assembly_fragment"
133298..171763
/note="assembly_fragment"
BASE COUNT 48403 a 34526 c 33542 g 53768 t 1524 others
ORIGIN

Query Match 5.3%; Score 90.2; DB 2; Length 171763;
Best Local Similarity 56.1%; Pred. No. 1.6e-06;
Matches 170; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 194 AAAAAAGAGAAAATAAGATGAGAGATAGAGATTTCTATGTAAGAAAAGAGAGAGAAC 253
Db 130691 AGAGAGAGGGGGAGGGAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 130632
QY 254 ATGTAGGTGAACAAAATAAAGAGATATGATATATTTTATGAGAGTGTGAAGATTA 313
Db 130631 AAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAGAGAGAGAGAGAGA 130572
QY 314 TTTTAGGAGAGGAGAGAGAAATGAAAAGAAAATGACATGGTGAATCTGAAGAAGATG 373
Db 130571 GAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAGAGAGAGA 130512
QY 374 AATTGTGTTAAGATGAAGAGAGAAAGAGAACTCCATGGCTAAAGTCTCGTAAGAGAT 433
Db 130511 AAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAGAGAGAGAGAGA 130452
QY 434 GAAAAAGAAACAAAAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 493

Db 130451 AAGAAAGA 130392
QY 494 AAA 496
Db 130391 AAA 130389
RESULT 15
AC125184/c
LOCUS
DEFINITION
280982 bp DNA linear HTG 23-JUN-2002
Mus musculus chromosome UNK clone RP23-29; 59, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
AC125184
VERSION
AC125184.1 GI:21490707
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 280982)
McPherson, J.D. and Waterston, R.H.
TITLE
The sequence of Mus musculus clone
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 280982)
McPherson, J.D. and Waterston, R.H.
TITLE
Direct Submission
JOURNAL
Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
3 (bases 1 to 280982)
McPherson, J.D. and Waterston, R.H.
TITLE
Direct Submission
JOURNAL
Submitted (23-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0299L09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990315
Consensus quality: 277729 bases at least Q40
Consensus quality: 279043 bases at least Q30
Insert size: 197000; agarose-fp
Insert size: 281848; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 9.94 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8161: contig of 8161 bp in length
* 8162 8261: gap of unknown length
* 8262 51021: contig of 42760 bp in length
* 51022 51122: gap of unknown length
* 51122 84880: contig of 33759 bp in length
* 84881 84981: gap of unknown length
* 84981 280982: contig of 196002 bp in length.
Location/Qualifiers

FEATURES

```
source      1. .280982
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="UNK"
            /clone="RP23-299L9"
            1. .8161
misc_feature /note="assembly_name:Contig10"
8262..51021
misc_feature /note="assembly_name:Contig11"
51122..84880
misc_feature /note="assembly_name:Contig12"
84981..280982
misc_feature /note="assembly_name:Contig13"
90913 a 51036 c 51472 g 87256 t 305 others
BASE COUNT
ORIGIN
Query Match      5.2%; Score 89.2; DB 2; Length 280982;
Best Local Similarity 54.6%; Pred. No. 2.2e-06;
Matches 178; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 138 TTTTAAATCAATCTATCTCTTATATACACACGACATTGACGGAAATTCAGGTAAAA 197
Db 162899 TTTTAACTGAAAAAAGCAAAACCTCATCAAAAAAGAAAGAAAGCAAGAAAGAGAGAGAGA 162840
QY 198 AGAGAAAAATAAAGATGAGAGATAGAGAGATTCTTATGCAAAAAAGAAAGAGAACATGT 257
Db 162839 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162780
QY 258 AGGTGAACAAATAAAGAGATATGATATATTTTATGAGAGGTGGTGAAGATTTT 317
Db 162779 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162720
QY 318 AGGAGAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 377
Db 162719 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162660
QY 378 GTGTTAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 437
Db 162659 GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 162600
QY 438 AAGAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
Db 162599 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 162574
Search completed: November 29, 2002, 00:40:31
Job time : 5728 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 22:59:21 ; Search time 2040 Seconds
(without alignments)
13496.239 Million cell updates/sec

Title: US-09-733-685-3

Perfect score: 1700

Sequence: 1 tgggtttttattgataaca.....tttacgacttacgaacatat 1700

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230	13.5	230	17	BH866477
2	203.8	12.0	531	17	BH696479
3	201.2	11.8	206	17	BH866474
4	185	10.9	237	17	BH814587
5	175.2	10.3	196	17	BH866500
6	173.6	10.2	267	17	BH854168

C	7	128.4	7.6	387	17	BH749275
C	8	101.6	6.0	1101	17	CNS0021J
C	9	96	5.6	660	17	BH183498
C	10	96	5.6	660	17	CNS070NJ
C	11	91	5.4	982	17	AQ325799
C	12	89	5.2	529	9	AL514657
C	13	88	5.2	768	17	CNS01V5E
C	14	86.6	5.1	1101	17	CNS008VL
C	15	86.4	5.1	764	17	AZ193158
C	16	85.8	5.0	842	17	CNS010QN
C	17	85.6	5.0	997	17	CNS005TE
C	18	85	5.0	886	17	CNS075DF
C	19	85	5.0	934	17	AZ184244
C	20	85	5.0	1201	17	CNS0167M
C	21	84.8	5.0	1101	17	CNS00LT2
C	22	84	4.9	647	17	BH038922
C	23	84	4.9	839	14	BQ151187
C	24	83.8	4.9	922	17	AZ548363
C	25	83.4	4.9	638	17	AQ329262
C	26	83.4	4.9	848	17	AZ549808
C	27	82.6	4.9	89	17	BH866492
C	28	82.6	4.9	718	17	AG160666
C	29	82.4	4.8	434	17	CNS060Z3
C	30	82.4	4.8	952	17	AQ897460
C	31	82.2	4.8	580	17	CNS008HN
C	32	82.2	4.8	1101	17	CNS0039G
C	33	82	4.8	888	17	AZ549422
C	34	81.8	4.8	625	17	CNS036A2
C	35	81.8	4.8	815	17	AG044049
C	36	81.4	4.8	891	17	CNS009JU
C	37	81.2	4.8	1001	17	CNS0064G
C	38	81.2	4.8	1084	17	CNS06PLB
C	39	81.2	4.8	1101	17	CNS012TP
C	40	81	4.8	89	17	BH866501
C	41	80.8	4.8	1042	17	CNS0148K
C	42	80.6	4.7	797	10	BE034282
C	43	80.6	4.7	1101	17	CNS0181N
C	44	80.4	4.7	614	17	CNS0152H
C	45	79.8	4.7	987	17	CNS00418

ALIGNMENTS

RESULT 1
BH866477 230 bp DNA linear GSS 05-AUG-2002
LOCUS SALK_101380 Arabidopsis thaliana TDNA insertion lines Arabidopsis
DEFINITION thaliana genomic clone SALK_101380, DNA sequence.
ACCESSION BH866477
VERSION BH866477.1 GI:22102375
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 230)

REFERENCE
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckersalk.edu
This is single pass sequence recovered from the left border of TDNA.

```

Class: TDNA tagged.
Location/Qualifiers
1. 230
/oranism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_I01380"
/clone.lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"

```

```

BASE COUNT      66 a      34 c      81 g      49 t
ORIGIN
Query Match      13.5%; Score 230; DB 17; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.1e-28;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 914 TCTGACGGTTAACAGAGGAAAACCGTTCGACGGAAGTTGTTCCGACGCGTAAACGGAGGAAGAG 973
      |||||
Db 1 TCTGACGGTTAACAGAGGAAAACCGTTCGACGGAAGTTGTTCCGACGCGTAAACGGAGGAAGAG 60
      |||||
QY 974 GTGCGATGAGTTTTTCAAGATATTACGGAGAGTACACGTGGCGACACGACGCGTTCGGAA 1033
      |||||
Db 61 GTGCGATGAGTTTTTCAAGATATTACGGAGAGTACACGTGGCGACACGACGCGTTCGGAA 120
      |||||
QY 1034 GTTAAACGGCGGTGTGCTGAGGGAGAGTTACCGTCTAAGAAGAGGAAACGGAGTCCAGAA 1093
      |||||
Db 121 GTTAAACGGCGGTGTGCTGAGGGAGAGTTACCGTCTAAGAAGAGGAAACGGAGTCCAGAA 180
      |||||
QY 1094 CTTGGGTTTGAGAAACTCCGTGGATTGTAACGGCGTTCCGACACGGAGAATT 1143
      |||||
Db 181 CTTGGGTTTGAGAAACTCCGTGGATTGTAACGGCGTTCCGACACGGAGAATT 230
      |||||

```

```

RESULT 2
BH696479/c
LOCUS
DEFINITION
531 bp DNA linear GSS 20-FEB-2002
BOHW78TR BO_2_3_KB Brassica oleracea genomic clone BOHW78, DNA
sequence.
ACCESSION
BH696479
VERSION
BH696479.1 GI:18769095
SOURCE
GSS.
ORGANISM
Brassica oleracea.
Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eutrosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 531)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOHW78TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubted haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

```

FEATURES	Location/Qualifiers	Source
	1 531	
	/organism="Brassica oleracea"	
	/strain="Tol000DH3"	
	/db_xref="taxon:3712"	
	/clone="BOHW78"	
	/clone_lib="BO_2_3_KB"	
	/note="Vector: pBOS1; Site 1: BstXI; 2-3 kb sheared"	

BASE COUNT 153 a 119 c 80 g 179 t

genomic DNA inserted into pHS1 using BstXI linkers"

ORIGIN

	Query Match	12.0%;	Score 203.8;	DB 17;	Length 531;
	Best Local Similarity	72.3%;	Pred. No. 1.7e-24;		
	Matches 297;	Conservative 0;	Mismatches 102;	Indels 12;	Gaps 2;
QY	950	GTTCGACCGTAAACGGAGAGAGAGGTGTGATGAGTTTTCAAGATATTACGGAGAGTACAC	1009		
DB	531	GTACGTACCGTCACGGAGGAGAGGTGGATGAGTTCTCAAGATATTACGGAGACTACAC	472		
QY	1010	GTGGGCACACGACCGGTTGCCAAGATTAAACGGCGGTTGCTCAGGCGAGAGTACCGTCT	1069		
DB	471	GTGGCGACACGGATTGTGGCGAGAGTTAAACGGCGGTGACCTGAAGGTGATAGCTTCT	412		
QY	1070	AAGAAGAGAAACGGAGTCAGAAATCTTGGGTTGAGAACTCGTTGGATTGTAAACGGCTT	1129		
DB	411	AAGAAGAGAAACGGATTACAGTCTTGGGCTGAGGAGCTCATTTGGATACTAACGAAGTT	352		

Qy	1130	CGAGACGGAGAAATTCGATCAGATAATATCGGTCGGGTTACAGGGTTGGTGGTTCCGATCTG	1189
Db	351	CAGATGGGAAATCAGATGGAAATAAATCGGTCGGGTACGGAATTTGGAGTCTT	292
Qy	1190	AACTGTAAACCGGAACGACAGACGGCTAGTTATTCGTTGTAGACTTGTAGTCCTTCATGT	1249
Db	291	AACTGTCAACGGGAACCCGAGCAGTTA-----AAATGTATCTCTTCTCAATA	243
Qy	1250	TTTTCCCTCTTTACAATAATCAATTTT-TTTTTAACTAATACCTTTTCAAAAATAATGG	1308
Db	242	CTAAATATTGTTCCAGTAATCTATTTTCAATATACTAAATATTTTATTATTAGCTGG	183
Qy	1309	TAAAGAAGATTAATTAACATGTTATCCAAATTTTCAGATTCCTCAGTTTAT	1359
Db	182	TTCTTAAAAACGCTACTTAAATATTTATCCAGTTTCAGATTCGCTCTTTGATTTT32	

RESULT_3				
BH866474				
LOCUS				
DEFINITION				
	BH866474	206 bp	DNA	linear GSS 05-AUG-2002
	SALK_101377 Arabidopsis thaliana TDNA insertion lines Arabidopsis			
	thaliana genomic clone SALK_101377. DNA sequence.			

accession BH866474
 version BH866474.1
 keywords GI:22102372
 source GSS.
 organism
 thale cress.
 Arabidopsis thaliana
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 206)
 Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
 ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
 , Zimmerman,J., and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 Unpublished (2001)
 Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA.
 Class: TDNA Tagged.

```

FEATURES
source
1. .206
Location/Qualifiers
Class: lona tagged:
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clones="SALK 101377"

```


Matches 183; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 948 TTGTTCCGACGGTAACGGAGGAAGAGGTGGATGACTTTTCAAGATATTACGGAGACTAC 1007

Db 1 TTGTTCCGACTGCTCGGGGGAGAGGTCGATGACTTTTCAAGATATTACGGAGACTAC 60

Qy	1008	ACGTGGGACACGAACGGTTGC	AAAGTTAACGGCGGTGCTG	AGGGAGACTTACCGT	1067
Db	61	ATTGGCGACCCGAACGGTTC	GACAGTTTAACGGCGGTGTT	GCTGAAGGAGACTTACCGT	120

Qy	1068	CTAAGACAGAGGAACGGGACTCAGAATCTTTGGGTTCAGAAACTCGTGGATTGTAAACGCCG	1127
Db	121	CTATGAACAGAGGAACGGGACCAGAATCTTTGGGTTTCAGAAACTCGTGGATTGTAAACGCCG	180

Qy 1128 TTCGAGACGGAGAATT 1143
|||||
Db 181 TTCGAGACGGAGAATT 196

RESULT 6	BH854168	267 bp	DNA	linear	GSS 13-JUN-2002
LOCUS	BH854168				
DEFINITION	SALK_078835..55.00.x Arabidopsis thaliana TDNA insertion lines				
	Arabidopsis thaliana genomic clone SALK_078835..55.00.x, DNA sequence.				

ACCESSION	BH854168
VERSION	BH854168.1
KEYWORDS	GSS,
SOURCE	thale grass.

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidoneae

REFERENCE
1 (bases 1 to 267)
AUTHORS
Alonso, J. M., Leisse, T. J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C. J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Foster, J. D.

TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome Unpublished (2001)
JOURNAL	
COMMENT	Contact: Yousang Park

Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 452 4100 ext. 3758

Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of

```

FEATURES
  source      1. .267
              Location/Qualifiers
Class: TDNA tagged.

```

```

/organism="Aricaduopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_078835.55.00.x"

```

```

/clonelib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was

```

directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html.

ORIGIN

Query Match 10.2%; Score 173.6; DB 17; Length 267;

	Matches	179;	Conservative	0;	Mismatches	9;	Indels	0;	Gaps	0;
QY	1	TGGGTTTTTT	ATTCGATAACATGACAAATATTTATTTATTTTCATGACGTTTTTTATTTGGATAG	60						

Db 80 TGAGTTTTTATTTGGATAACATGACAAATATTTATATACATGAATTTTCTTGGATAG 139

Qy 61 CATGACAAATATTAAATATATATCAGTGTAAATAACATGTTTGTCTTAAATACATGCATT 120
 |||||
 Db 140 CATGACAAATATTAAATATATCAGTGTAAATAACATGTTTGTCTTAAATACATGCATT 199

Qy	121	TTAAATCAGACATTGTTTAAATCAAAATCTAATCTTTATATCACAAAGACATTGAC	180
Db	200	TTAAATCAGACATTGTTTAAATCAAAATATAATCTTTATATCACAAAGACATTGAT	259

Qy 181 GGAAATT 188
| | | | |
Db 260 GAAGAATT 267

RESULT 7
BH749275/C

LOCUS BH749275 387 bp DNA linear GSS 27-FEB-98
DEFINITION SALK_047543.32.95.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_047543.32.95.x, DNA sequence

ACCESSION	BH749275
VERSION	BH749275.1
KEYWORDS	GI:18963963 GSS.

ORGANISM *Arabidopsis thaliana*
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
1 (bases 1 to 387)
Alonso, J. M., Leisse, T. J., Barajas, P., Chen, H., Cheuk, R., Gadri, C., Jeske, A., Karnes, M., Kim, C. J., Parker, H., Prednis, L., Shi

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)

CONTACT: JOSEPH N. DEKRE
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

tel: 858 433 4100 x1732
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of

TDNA.	Class: TDNA tagged.
FEATURES	Location/Qualifiers
source	1 387

```

/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/ncbi_taxonomy="catal:047543.22.05.4"

```

```
/clone_lib="Arabidopsis thaliana TDNA insertion lines
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
```

directly sequenced to determine the genomic sequence the site of insertion. Details of the protocols used be found at http://signal.salk.edu/tdna_protocols.htm

Query Match 7.6%; Score 128.4; DB 17; Length 387;

Qv 185 AATTCAGCTAAAAACAGAAAAATAAAGAATGAGAGATAGAGAGAGATTTCTATTGGAAGAAAGAA 244
Matches 149; Conservative 0; Mismatches 16; Indels 1; Gaps
Best Local Similarity 89.8%; Pred. NO. 6.7e-12;

Db 344 AATTCAGCTAAAAGAGAAAATAAGAAATGAGAGATAGAGAGATTCTATGGAAAATGAA 285

Q4 245 ACACACACACATCTACCTCCGAC - AAAATAAACAACACATATACATCATATATATATACACACCTC 202

Db 284 ACAGAGAACATGTGGGTGATCAAAATAAAGAGATATCATGATATATTTATGAGAGGTG 225

[illegible]

AQ325799/c
 LOCUS nbx0021B14r 982 bp DNA linear GSS 08-JAN-1999
 DEFINITION nbx0021B14r CUGI Rice BAC Library Oryza sativa genomic clone
 ACCESSION AQ325799
 VERSION AQ325799.1 GI:4117649
 KEYWORDS GSS.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 1 (bases 1 to 982)
 WING, R.A. and Dean, R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: GGAAACAGCTATGACCATG
 Class: BAC ends
 High quality sequence start: 4
 High quality sequence stop: 123.
 Location/Qualifiers
 1..982
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nbx0021B14r"
 /clone_lib="CUGI Rice BAC Library"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
 HindIII; Rice is one of two most popular grains in the
 world. Half of the world population especially those
 inhabiting highly populated areas of the humid tropics
 and subtropics, rely on rice as their primary source of
 carbohydrate. Monocotyledonous rice is a diploid plant
 (2n=24) with a haploid genome equivalent of 431 Mbp
 (Arumuganathan and Earle, 1991). The relatively small
 genome of rice, three times larger than that of
 Arabidopsis, makes it suitable for genomic studies. In
 order to facilitate positional cloning, physical mapping
 and genome sequencing of rice, we have constructed a BAC
 library from Oryza sativa, Nipponbare variety. The
 library contains 36,864 clones with an average insert size
 of 128.5 Kb providing 10.9 haploid genome equivalents. The
 deep coverage allows the isolation a particular sequence
 with a probability of 99.9 %. Two high density filters,
 each containing 18,432 clones (doubly spotted), represent
 the whole library for colony screening."
 141 a 69 c 43 g 674 t 55 others

FEATURES
 source

BASE COUNT 141 a 69 c 43 g 674 t 55 others
 ORIGIN
 Query Match 5.4%; Score 91; DB 17; Length 982;
 Best Local Similarity 46.9%; Pred. No. 7.7e-06;
 Matches 229; Conservative 0; Mismatches 259; Indels 0; Gaps 0;
 QY 9 TATTGGATACATGACAAATATTATTATTTCATGAGCTTTTATTTGGATAGCATGACAA 68
 |||||
 DB 903 TATGGAATATAGAAAAGATTTATTTATATAAATAATTTTATTTATTTAATAAT 844
 QY 69 ATATTAAATATACATGTTTAAATACATGTTTCTTTAAATACATGATTTTAAATC 128
 |||||
 DB 843 TGATTAATATATTAATAATATATATTTTATAAAAATAATAATTAATTAATTAAT 784
 QY 129 ACACATTTGTTTAAATCAATCTAATCTTATATACACACATGACGGAATTT 188
 |||||

Db 783 AATATNTATTGGATAATTAATATATATATATTTTAAATATATATAATAATAAATA 724
 QY 189 CAGGTAAAGAGAGAAATAAAGATAGAGATAGAGAGATTTCTATGCAAAAAGAGAG 248
 |||||
 Db 723 AAAAAAANNNNAANNAATTAATAAANNAANNAANNAANNAANNAANNAANNA 664
 QY 249 AGACATGTAGTGAACAAATAAAGAGATATGATGATATATTTTATGAGAGGCTG 308
 |||||
 Db 663 AANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 604
 QY 309 GATTATTTTAGGAGGAGAGAGAAATAGAAAAGAAATGACATGCTGAATCTGAAGA 368
 |||||
 Db 603 AANATATAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 544
 QY 369 AGATGAATTTGTTTAAAGATGAAGAGAGAGAGAAATCCATGCTGCTGCTGCTG 428
 |||||
 Db 543 AAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 484
 QY 429 AAGATGAAG 488
 |||||
 Db 483 AAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 424
 QY 489 TCCCAAAA 496
 |||||
 Db 423 AAAAAAAA 416
 RESULT 12
 LOCUS AL514657
 DEFINITION AL514657 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BB0022E09 3
 prime, mRNA sequence.
 ACCESSION AL514657
 VERSION AL514657.1 GI:12778151
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 529)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 JOURNAL
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CL0BB0022E09"
 /clone_lib="LTI_NFL006_PL2"
 /tissue_type="placenta"
 /notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technology,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

FEATURES
 source

BASE COUNT 364 a 20 c 9 g 89 t 47 others
 ORIGIN
 Query Match 5.2%; Score 89; DB 9; Length 529;
 Best Local Similarity 43.9%; Pred. No. 2.1e-05;
 Matches 206; Conservative 35; Mismatches 228; Indels 0; Gaps 0;
 QY 28 TATTATTTTATTCATGAGTTTTTATTTGGATAGCATGACAAATATTAAATATAC 87
 |||||

OY 485 CTATTGCCAAAA 496
|
|
|
|
Db 248 AAAAAAAAAAAAA 237

Search completed: November 29, 2002, 00:31:28
Job time : 2058 secs


```

QY 227 ATTTCTATGGAAAGAAAGAGAGAACATGTAGTGTAACAAATAAAGAGATATGATGAT 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 AAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 287 ATATTTATGAGAGGTGGTGAAGATTATTTAGAGAGGGAGAGAGAAATAGAAAAGAA 346
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 294 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 347 AATGACATGGTGAATCTGAAGAAGATGAATTTGTTAAAGATGAAGAGAGAAAGAGAACT 406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 407 CCATGGCTAAATCTCGCTAAAGAGATGAAAGAAAGAAAGAAAGAAAGAAAGAGAG 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 467 AAGGCTAAATAGACTTAACCTATTCGCCAAA 496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-09-960-352-3400/C
; Sequence 3400, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Tao, Nengbing
; APPLICANT: Bvatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3400
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 15-LIB3058-052-Q1-K1-D11
US-09-960-352-3400

Query Match 4.2%; Score 72; DB 10; Length 446;
Best Local Similarity 48.5%; Pred. No. 1.4e-05;
Matches 198; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 108 AAATACATGCAATTTAAATCAGACATTTGTTTAAATCAATCTAATCTTATATCA 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 168 CAACGACATTCAGCGAAATTCAGGTAAAGAGAAATAAAGAAATAGAGATAGAGAGA 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 228 TTTCTATGGAAAAAGAGAGAACATGTAGTGTAACAAATAAAGAGATATGATGATA 287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 AAATAAATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAA 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 288 TATTTATGAGAGGTGGTGAAGATTATTTAGGAGAGGGAGAGAGAAATAGAAAAGAAA 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 AATATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 348 ATGACATGGTGAATCTGAAGAGATGAATTTGTTAAAGATGAAGAGAGAAAGAGAACTC 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 408 CATGCTTAAGCTCTGTAAGAACATGAAAGAAAGAAAGAAAGAAAGAAAGAGAAAGAG 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 50
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 468 AGGCTAAATAGACTTAATTCGCAAAATTTCTGTAGCGCAACAATA 515
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 49 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGCGCGCTCGCATCTA 2

RESULT 3
US-10-095-407-16/C
; Sequence 16, Application US/10095407
; Patent No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; CURRENT FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-16

Query Match 4.2%; Score 71.2; DB 9; Length 152331;
Best Local Similarity 54.6%; Pred. No. 9.5e-05;
Matches 142; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 203 AAATAAAGAAATGAGATAGAGAGATTTCTATGGAAAAAGAGAGAAACATGTAGGTG 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128751 AGACAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 263 AACAAAAATAAGAGATATGATGATATATTTATGAGAGGTGGTGAAGATTATTTTAGGAG 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128691 AGAAAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 323 AGGGAGAGAGAAATAGAAAAAGAAAAATGACATGTTGTAATCTGAAGAAAGATGAATTTCTTT 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128631 AGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 383 AAAGATGAAGAGAGAGAGAGAACTCCATGGCTTAAAGTCTCGTAAAGAAAGATGAAAAAGAA 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128571 AAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 443 ACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128511 AGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128492
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-10-095-407-17/C
; Sequence 17, Application US/10095407
; Patent No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; CURRENT FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17

```



```

; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo
; FEATURE:
; NAME/KEY: misc.
; LOCATION: (1).
; OTHER INFORMAT
US-10-095-407-17

```

Query Match	4.28;	Score 71.2;	DB 9;	Length 176373;
Best Local Similarity	54.68;	Pred. No. 9.9e-05;		
Matches 142;	Conservative 0;	Mismatches 118;	Indels 0;	Gaps 0;
QY 203	AAATAAAGAAATGAGAGATAGAGAGATTTCTATGGAAGAAAGAAAGAGAGAACATGTTAGGTG	262		
Db 144774	AGACAAGAAAGAAAGAGAGAGAGGGGAGGAGAAAGAAAGAGAGAAAGAGAAAGAAAGAA	144715		
QY 263	ACAAATAAAGAGATATGATCATATATTTATGAGAGGTGGTGAAGATTATTTAGGAG	322		
Db 144714	AGAAAAGAAAGAAAGAGAGAGAGGGGAGGGAGAGAGAAAGAAAGAAAGAGAAAGAG	144655		
QY 323	AGGGAGAGAGAAATAGAAAAAGAAATGACATGTGTAATCTGAAGAGATGTAATTTGTGTT	382		
Db 144654	AGAAAGAGAGAAAGAGAAAGAAAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	144595		
QY 383	AAAGATGAAGAGAAAGAGAGACTCCATGGCTAAAGTCTCGTGAAGAGATGAAAAGAA	442		
Db 144594	AAAGAAGAAAGAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAAAGAAAGAAAGAGAGGA	144535		
QY 443	ACAAAAGAAAGGAAGAGAA	462		
Db 144534	AGGAAGGAAGGAAGGAAGAA	144515		

RESULT 5
US-09-867-701-9583

```

; Sequence 9583: Application US/09867701
; Patent No.: US20020132237A1
; GENERAL INFORMATION:
;
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9583
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9583

```

Query Match	4.0%	Score 68	DB 10	Length 292
Best Local Similarity	52.9%	Pred. No. 7.6e-05		
Matches 146	Conservative 0	Mismatches 130	Indels 0	Gaps 0
QY 196	AAAGAGAAATTAAGAATGAGAGATAGAGAGATTCTTATGGAATAAAGAGAGAGAACAT	255		
Db				
10	ACAGAGGGGAGGCG	69		
QY 256	GTAGGTGAACAAATAAGACATATGATATATATTTATGAGAGCTGCTGAAGATTATT	315		
Db				
70	GAGCGGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCGGAGAGC	129		
QY 316	TTAGGAGGGGAGAGAGAAATAGAAAAGAAATGACATGGTGAATCTCTGAAGAGATGAA	375		
Db				
130	GAGAGACAGAGAGAGAGAGAGAGAGAGAAACCGCGAGGAGAGAGCGGAGAGAGAGAGA	189		
QY 376	TTGTGTTAAAGATGAAGAGAGAAAGAGAAATCTCATGGCTAAAGTCTCGTAAAGAGAGATGA	435		
Db				

```

Db 190 AAACGACGGAGGGAGGAGCGAGCGAGCGAGGAGAAAGAAAGAAAGAAAGAA 2499
Qy 436 AAAAGAAACAAGAGGAGGAGAGAGAGAGAGAGAGAGAGGC 471
      | | | | | | | | | | | | | | | | | | | | | |
Db 250 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTC 285

RESULT 6
US-09-960-352-5785/C
; Sequence 5785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LPA
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5785
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (76),(90)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
; JS-09-960-352-5785

```

Query Match
4.0%; Score 68; DB 10; Length 516;

[illegible]

GENERAL INFORMATION:
APPLICANT: Haglaczkv, Gyula
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS
FOR PREPARING ARTIFICIAL CHROMOSOMES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSER: Heller Ehrman White & McAuliffe LLP
STREET: 4350 La Jolla Village Drive, 7th Floor
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/125,767
FILING DATE: 17-Apr-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/724,693
FILING DATE: 28-NOV-2000
APPLICATION NUMBER: 08/835,682
FILING DATE: 10-APR-1997
APPLICATION NUMBER: 08/695,191
FILING DATE: 07-AUG-1996
APPLICATION NUMBER: 08/682,080
FILING DATE: 15-JUL-1996
APPLICATION NUMBER: 08/629,822
FILING DATE: 10-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24601-402J
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858-450-8403
TELEFAX: 858-587-5360
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 42999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: <Unknown>
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-125-767-17
Query Match 3.9%; Score 65.6; DB 9; Length 42999;
Best Local Similarity 52.5%; Pred. No. 0.00084;
Matches 166; Conservative 0; Mismatches 149; Indels 1; Gaps 1;
Qy 166 CACACACATGCGGAAATTCAGGTAAAGAGAGAAATTAAGATGAGATAGAGA 225
Db 27215 CAGACAGACGGGGAGGAGGAGACAAACAGAGAGAGAGAGAGAGAGAGAGA 27156
Qy 226 GATTTCATGAAAAAGAGAGACATGTAGGTGAACAAAATAAAGAGATATGATGA 285
Db 27155 GACAGACAGACAGACAGACAGACAAAGAACAGACAGAAAGAGAGAGACGGA 27096
Qy 286 TATATTTTATCAGAGGTGGTGACATTTTATAGGAGGGGAGAGAGAAATAGAAAAAGA 345
Db 27095 GAGAGAGTGAAGAGGGAGAGAGACATGGAGGGGAGAGACAGACAGAGAGAGA 27036
Qy 346 AAATGACATGGTGAATCTGAAGAGATGAATTCGTGAAGATGAACAGAGAAAGAGAAC 405
Db 27035 AACAGACAGAAAGAGAGAGACGCGGAGAGAGATGAGTGAGAGAGAGAGAGAGACA 26976

Qy 406 TCCATGGCTAAAGTCTCGTTAAAGA-AGATGAAAAAGAAACAAAGAGAGAGAAAGA 464
Db 26975 TGGAGGAGAGACAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGACGAGAGAGA 26916
Qy 465 GAAAGGCTAAATAGA 480
Db 26915 GTGAGTGAGAGAGAGA 26900
RESULT 13
US-09-960-352-10262/c
; Sequence 10262, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10262
; LENGTH: 451
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 44-LIB3058-031-Q1-K1-C8
US-09-960-352-10262
Query Match 3.8%; Score 65.2; DB 10; Length 451;
Best Local Similarity 51.7%; Pred. No. 0.0003;
Matches 148; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
Qy 183 AAATTCAGGTAAAGAGAGAAATTAAGCATGAGAGATAGAGAGATTCTATCGAAAAAG 242
Db 438 AAAAGAGAAAAAGAAAAAGAAAAATTAAGAAAAAGAAAAAGTAAAAAGAAAAAGAG 379
Qy 243 AAAGAGAGAACATGTAGGTGAACAAATTAAGAGATATGATGATATATTTATGAGAGGT 302
Db 378 AAAAGAAAAAGAAAAATTAAGAAAAAGAAAAATTAAGAAAAAGAAAAAGAAAAAG 319
Qy 303 GGTGAAGATTATTTAGGAGGGGAGAGAGAAATAGAAAAAGAAAAATGACATGCTGAATC 362
Db 318 AAATTAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 259
Qy 363 TGAAGAAGATGAATTGTTAAAGATGAGAGAGAGAGAACTCCATGGCTAAGTCTC 422
Db 258 AAAAGAGAAAAATTAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 199
Qy 423 GTAAAGAGAGATGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 468
Db 198 AAATTAAGAAAAAGAGAGATTAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 153
RESULT 14
US-09-967-768A-314/c
; Sequence 314, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111

